

1. [Preface](#)
2. Welcome to Economics!
  1. [Introduction](#)
  2. [What Economics Is and Why It's Important](#)
  3. [Microeconomics and Macroeconomics](#)
  4. [How Economists Use Theories and Models to Understand Economic Issues](#)
  5. [How Economies Can Be Organized: An Overview of Economic Systems](#)
3. Choice in a World of Scarcity
  1. [Introduction to Choice in a World of Scarcity](#)
  2. [How Individuals Make Choices Based on Their Budget Constraint](#)
4. Demand and Supply
  1. [Introduction to Demand and Supply](#)
  2. [Demand, Supply, and Equilibrium in Markets for Goods and Services](#)
  3. [Shifts in Demand and Supply for Goods and Services](#)
  4. [Changes in Equilibrium Price and Quantity: The Four-Step Process](#)
  5. [Price Ceilings and Price Floors](#)
  6. [Demand, Supply, and Efficiency](#)
5. Consumer Choices
  1. [Introduction to Consumer Choices](#)
  2. [Consumption Choices](#)
  3. [How Changes in Income and Prices Affect Consumption Choices](#)
  4. [Labor-Leisure Choices](#)
  5. [Intertemporal Choices in Financial Capital Markets](#)
6. Cost and Industry Structure
  1. [Introduction to Cost and Industry Structure](#)

2. [Explicit and Implicit Costs, and Accounting and Economic Profit](#)
  3. [The Structure of Costs in the Short Run](#)
  4. [The Structure of Costs in the Long Run](#)
7. Perfect Competition
  1. [Introduction to Perfect Competition](#)
  2. [Perfect Competition and Why It Matters](#)
  3. [How Perfectly Competitive Firms Make Output Decisions](#)
  4. [Entry and Exit Decisions in the Long Run](#)
  5. [Efficiency in Perfectly Competitive Markets](#)
8. Monopoly
  1. [Introduction to a Monopoly](#)
  2. [How Monopolies Form: Barriers to Entry](#)
  3. [How a Profit-Maximizing Monopoly Chooses Output and Price](#)
9. Monopolistic Competition and Oligopoly
  1. [Introduction to Monopolistic Competition and Oligopoly](#)
  2. [Monopolistic Competition](#)
  3. [Oligopoly](#)
10. Monopoly and Antitrust Policy
  1. [Introduction to Monopoly and Antitrust Policy](#)
  2. [Corporate Mergers](#)
  3. [Regulating Anticompetitive Behavior](#)
  4. [Regulating Natural Monopolies](#)
  5. [The Great Deregulation Experiment](#)
11. [The Use of Mathematics in Principles of Economics](#)
12. [Indifference Curves](#)
13. [Present Discounted Value](#)

## Preface

*Principles of Microeconomics* is designed for a one-semester microeconomics introductory course. It is traditional in coverage, including introductory economics content, microeconomics, and international economics. At the same time, the book includes a number of innovative and interactive features designed to enhance student learning. Instructors can also customize the book, adapting it to the approach that works best in their classroom.

Welcome to *Principles of Microeconomics*, an OpenStax resource. This textbook has been created with several goals in mind: accessibility, customization, and student engagement—all while encouraging students toward high levels of academic scholarship. Instructors and students alike will find that this textbook offers a strong foundation in microeconomics in an accessible format.

## About OpenStax

OpenStax is a non-profit organization committed to improving student access to quality learning materials. Our free textbooks go through a rigorous editorial publishing process. Our texts are developed and peer-reviewed by educators to ensure they are readable, accurate, and meet the scope and sequence requirements of today's college courses. Unlike traditional textbooks, OpenStax resources live online and are owned by the community of educators using them. Through our partnerships with companies and foundations committed to reducing costs for students, OpenStax is working to improve access to higher education for all. OpenStax is an initiative of Rice University and is made possible through the generous support of several philanthropic foundations.

## About OpenStax's Resources

OpenStax resources provide quality academic instruction. Three key features set our materials apart from others: they can be customized by instructors for each class, they are a "living" resource that grows online through contributions from science educators, and they are available free or for minimal cost.

## Customization

OpenStax learning resources are designed to be customized for each course. Our textbooks provide a solid foundation on which instructors can build, and our resources are conceived and written with flexibility in mind. Instructors can select the sections most relevant to their curricula and create a textbook that speaks directly to the needs of their classes and student body. Teachers are encouraged to expand on existing examples by adding unique context via geographically localized applications and topical connections.

*Principles of Microeconomics* can be easily customized using our online platform (<http://cnx.org/content/col11627/>). Simply select the content most relevant to your current semester and create a textbook that speaks directly to the needs of your class. *Principles of Microeconomics* is organized as a collection of sections that can be rearranged, modified, and enhanced through localized examples or to incorporate a specific theme of your course. This customization feature will ensure that your textbook truly reflects the goals of your course.

## Curation

To broaden access and encourage community curation, *Principles of Microeconomics* is “open source” licensed under a Creative Commons Attribution (CC-BY) license. The economics community is invited to submit examples, emerging research, and other feedback to enhance and strengthen the material and keep it current and relevant for today’s students. Submit your suggestions to [info@openstaxcollege.org](mailto:info@openstaxcollege.org).

## Cost

Our textbooks are available for free online, and in low-cost print and e-book editions.

## About *Principles of Microeconomics*

*Principles of Microeconomics* has been developed to meet the scope and sequence of most introductory microeconomics courses. At the same time, the book includes a number of innovative features designed to enhance student learning. Instructors can also customize the book, adapting it to the approach that works best in their classroom.

## Coverage and Scope

To develop *Principles of Microeconomics*, we acquired the rights to Timothy Taylor's second edition of *Principles of Economics* and solicited ideas from economics instructors at all levels of higher education, from community colleges to Ph.D.-granting universities. They told us about their courses, students, challenges, resources, and how a textbook can best meet the needs of both instructors and students.

The result is a book that covers the breadth of economics topics and also provides the necessary depth to ensure the course is manageable for instructors and students alike. And to make it more applied, we have incorporated many current topics. We hope students will be interested to know just how far-reaching the recent recession was (and still is), for example, and why there is so much controversy even among economists over the Affordable Care Act (Obamacare). The Keystone Pipeline, Occupy Wall Street, and minimum wage debates are just a few of the other important topics covered.

The pedagogical choices, chapter arrangements, and learning objective fulfillment were developed and vetted with feedback from educators dedicated to the project. They thoroughly read the material and offered critical and detailed commentary. The outcome is a balanced approach to microeconomics, particularly to the theory and application of economics concepts. New 2015 data are incorporated for topics that range from average U.S. household consumption in Chapter 2 to the total value of all home equity in Chapter 17. Current events are treated in a politically-balanced way as well.

The book is organized into five main parts:

**What is Economics?** The first two chapters introduce students to the study of economics with a focus on making choices in a world of scarce resources.

**Supply and Demand**, Chapters 3 and 4, introduces and explains the first analytical model in economics: supply, demand, and equilibrium, before showing applications in the markets for labor and finance.

**The Fundamentals of Microeconomic Theory**, Chapters 5 through 10, begins the microeconomics portion of the text, presenting the theories of consumer behavior, production and costs, and the different models of market structure, including some simple game theory.

**Microeconomic Policy Issues**, Chapters 11 through 18, covers the range of topics in applied micro, framed around the concepts of public goods and positive and negative externalities. Students explore competition and antitrust policies, environmental problems, poverty, income inequality, and other labor market issues. The text also covers information, risk and financial markets, as well as public economy.

**International Economics**, Chapters 19 and 20, the final part of the text, introduces the international dimensions of economics, including international trade and protectionism.

Chapter 1 Welcome to Economics!

Chapter 2 Choice in a World of Scarcity

Chapter 3 Demand and Supply

Chapter 4 Labor and Financial Markets

Chapter 5 Elasticity

Chapter 6 Consumer Choices

Chapter 7 Cost and Industry Structure

Chapter 8 Perfect Competition

Chapter 9 Monopoly

Chapter 10 Monopolistic Competition and Oligopoly

Chapter 11 Monopoly and Antitrust Policy

Chapter 12 Environmental Protection and Negative Externalities

Chapter 13 Positive Externalities and Public Goods

Chapter 14 Poverty and Economic Inequality

Chapter 15 Issues in Labor Markets: Unions, Discrimination, Immigration

Chapter 16 Information, Risk, and Insurance

Chapter 17 Financial Markets

Chapter 18 Public Economy

Chapter 19 International Trade

Chapter 20 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics

Appendix B Indifference Curves

Appendix C Present Discounted Value

### **Alternate Sequencing**

*Principles of Economics* was conceived and written to fit a particular topical sequence, but it can be used flexibly to accommodate other course structures. One such potential structure, which will fit reasonably well with the textbook content, is provided. Please consider, however, that the chapters were not written to be completely independent, and that the proposed alternate sequence should be carefully considered for student preparation and textual consistency.

Chapter 1 Welcome to Economics!

Chapter 2 Choice in a World of Scarcity

Chapter 3 Demand and Supply

Chapter 4 Labor and Financial Markets

Chapter 5 Elasticity

Chapter 6 Consumer Choices

Chapter 19 International Trade

Chapter 7 Cost and Industry Structure

Chapter 12 Environmental Protection and Negative Externalities

Chapter 13 Positive Externalities and Public Goods

Chapter 8 Perfect Competition

Chapter 9 Monopoly

Chapter 10 Monopolistic Competition and Oligopoly

Chapter 11 Monopoly and Antitrust Policy

Chapter 14 Poverty and Economic Inequality

Chapter 15 Issues in Labor Markets: Unions, Discrimination, Immigration

Chapter 16 Information, Risk, and Insurance

Chapter 17 Financial Markets

Chapter 18 Public Economy  
Chapter 20 Globalization and Protectionism

Appendix A The Use of Mathematics in Principles of Economics  
Appendix B Indifference Curves  
Appendix C Present Discounted Value

## **Pedagogical Foundation**

Throughout the OpenStax version of *Principles of Microeconomics*, you will find new features that engage the students in economic inquiry by taking selected topics a step further. Our features include:

**Bring It Home:** This added feature is a brief case study, specific to each chapter, which connects the chapter's main topic to the real world. It is broken up into two parts: the first at the beginning of the chapter (in the Intro module) and the second at chapter's end, when students have learned what's necessary to understand the case and "bring home" the chapter's core concepts.

**Work It Out:** This added feature asks students to work through a generally analytical or computational problem, and guides them step-by-step to find out how its solution is derived.

**Clear It Up:** This boxed feature, which includes pre-existing features from Taylor's text, addresses common student misconceptions about the content. Clear It Ups are usually deeper explanations of something in the main body of the text. Each CIU starts with a question. The rest of the feature explains the answer.

**Link It Up:** This added feature is a very brief introduction to a website that is pertinent to students' understanding and enjoyment of the topic at hand.

## **Questions for Each Level of Learning**

The OpenStax version of *Principles of Microeconomics* further expands on Taylor’s original end of chapter materials by offering four types of end-of-module questions for students.

**Self-Checks:** Are analytical self-assessment questions that appear at the end of each module. They “click-to-reveal” an answer in the web view so students can check their understanding before moving on to the next module. Self-Check questions are not simple look-up questions. They push the student to think a bit beyond what is said in the text. Self-Check questions are designed for formative (rather than summative) assessment. The questions and answers are explained so that students feel like they are being walked through the problem.

**Review Questions:** Have been retained from Taylor’s version, and are simple recall questions from the chapter and are in open-response format (not multiple choice or true/false). The answers can be looked up in the text.

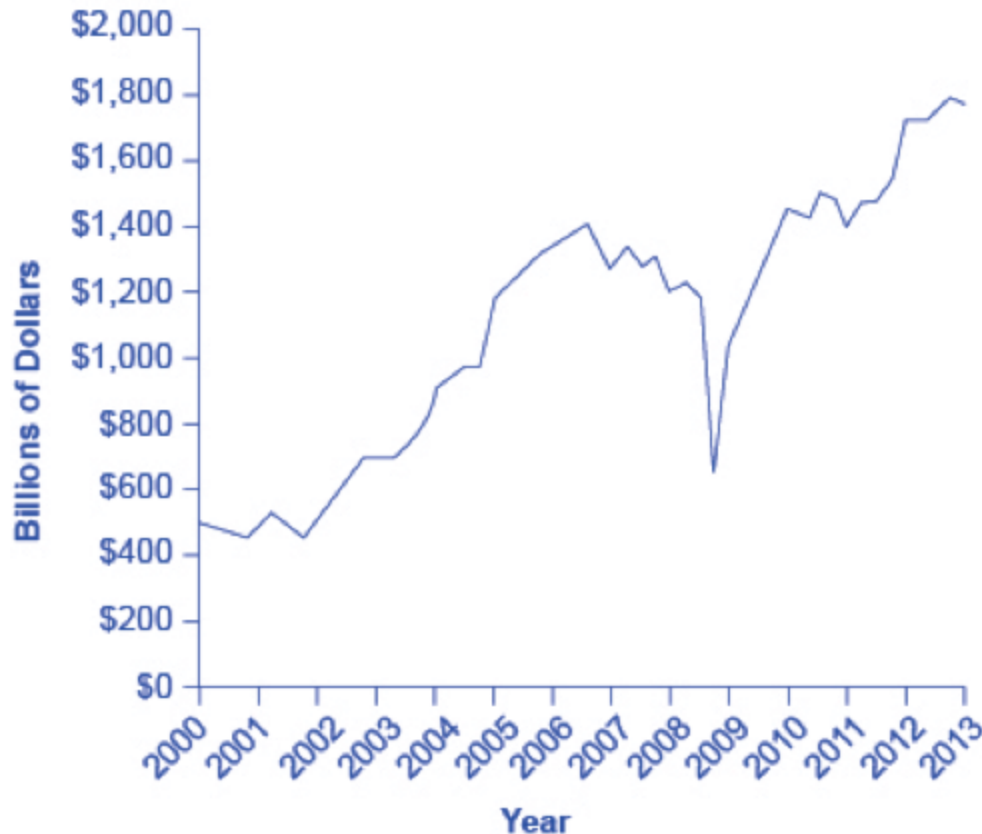
**Critical Thinking Questions:** Are new higher-level, conceptual questions that ask students to *demonstrate their understanding by applying* what they have learned in different contexts. They ask for outside-the-box thinking, for *reasoning* about the concepts. They push the student to places they wouldn’t have thought of going themselves.

**Problems:** Are exercises that give students additional practice working with the analytic and computational concepts in the module.

## Updated Art

*Principles of Microeconomics* includes an updated art program to better inform today’s student, providing the latest data on covered topics.

Corporate Profits After Tax (Adjusted for Inventory and Capital Consumption)



Since 2000, corporate profits after tax have mostly continued to increase each year, save for a substantial decrease between 2008 and 2009 as a result of the Great Recession. (Source: <http://research.stlouisfed.org/fred2>)

## About Our Team

### Senior Contributing Author

Timothy Taylor, Macalester College

Timothy Taylor has been writing and teaching about economics for 30 years, and is the Managing Editor of the *Journal of Economic Perspectives*, a post he's held since 1986. He has been a lecturer for The Teaching Company, the University of Minnesota, and the Hubert H. Humphrey

Institute of Public Affairs, where students voted him Teacher of the Year in 1997. His writings include numerous pieces for journals such as the *Milken Institute Review* and *The Public Interest*, and he has been an editor on many projects, most notably for the Brookings Institution and the World Bank, where he was Chief Outside Editor for the *World Development Report 1999/2000, Entering the 21st Century: The Changing Development Landscape*. He also blogs four to five times per week at <http://conversableeconomist.blogspot.com>. Timothy Taylor lives near Minneapolis with his wife Kimberley and their three children.

### **Senior Content Expert**

Steven A. Greenlaw, University of Mary Washington  
Steven Greenlaw has been teaching principles of economics for more than 30 years. In 1999, he received the Grellet C. Simpson Award for Excellence in Undergraduate Teaching at the University of Mary Washington. He is the author of *Doing Economics: A Guide to Doing and Understanding Economic Research*, as well as a variety of articles on economics pedagogy and instructional technology, published in the *Journal of Economic Education*, the *International Review of Economic Education*, and other outlets. He wrote the module on Quantitative Writing for *Starting Point: Teaching and Learning Economics*, the web portal on best practices in teaching economics. Steven Greenlaw lives in Alexandria, Virginia with his wife Kathy and their three children.

### **Senior Contributors**

Eric Dodge, Hanover College  
Cynthia Gamez, University of Texas at El Paso  
Andres Jauregui, Columbus State University  
Diane Keenan, Cerritos College  
Dan MacDonald, California State University San Bernardino  
Amyaz Moledina, The College of Wooster  
Craig Richardson, Winston-Salem State University

David Shapiro, Pennsylvania State University  
Ralph Sonenshine, American University

## **Reviewers**

Bryan Aguiar, Northwest Arkansas Community College  
Basil Al Hashimi, Mesa Community College  
Emil Berendt, Mount St. Mary's University  
Zena Buser, Adams State University  
Douglas Campbell, The University of Memphis  
Sanjukta Chaudhuri, University of Wisconsin - Eau Claire  
Xueyu Cheng, Alabama State University  
Robert Cunningham, Alma College  
Rosa Lea Danielson, College of DuPage  
Steven Deloach, Elon University  
Debbie Evercloud, University of Colorado Denver  
Sal Figueras, Hudson County Community College  
Reza Ghorashi, Richard Stockton College of New Jersey  
Robert Gillette, University of Kentucky  
George Jones, University of Wisconsin-Rock County  
Charles Kroncke, College of Mount St. Joseph  
Teresa Laughlin, Palomar Community College  
Carlos Liard-Muriente, Central Connecticut State University  
Heather Luea, Kansas State University  
William Mosher, Nashua Community College  
Michael Netta, Hudson County Community College  
Nick Noble, Miami University  
Joe Nowakowski, Muskingum University  
Shawn Osell, University of Wisconsin, Superior  
Mark Owens, Middle Tennessee State University  
Sonia Pereira, Barnard College  
Brian Peterson, Central College  
Jennifer Platania, Elon University  
Robert Rycroft, University of Mary Washington  
Adrienne Sachse, Florida State College at Jacksonville  
Hans Schumann, Texas AM

Gina Shamshak, Goucher College  
Chris Warburton, John Jay College of Criminal Justice, CUNY  
Mark Witte, Northwestern  
Chiou-nan Yeh, Alabama State University

## **Ancillaries**

OpenStax projects offer an array of ancillaries for students and instructors. Please visit <http://openstaxcollege.org> and view the learning resources for this title.

Introduction  
class="introduction"  
Do You Use Facebook?

Economics is  
greatly  
impacted by  
how well  
information  
travels through  
society. Today,  
social media  
giants Twitter,  
Facebook, and  
Instagram are  
major forces  
on the  
information  
super highway.  
(Credit: Johan  
Larsson/Flickr  
)

**Note:****Decisions ... Decisions in the Social Media Age**

To post or not to post? Every day we are faced with a myriad of decisions, from what to have for breakfast, to which route to take to class, to the more complex—“Should I double major and add possibly another semester of study to my education?” Our response to these choices depends on the information we have available at any given moment; information economists call “imperfect” because we rarely have all the data we need to make perfect decisions. Despite the lack of perfect information, we still make hundreds of decisions a day.

And now, we have another avenue in which to gather information—social media. Outlets like Facebook and Twitter are altering the process by which we make choices, how we spend our time, which movies we see, which products we buy, and more. How many of you chose a university without checking out its Facebook page or Twitter stream first for information and feedback?

As you will see in this course, what happens in economics is affected by how well and how fast information is disseminated through a society, such as how quickly information travels through Facebook. “Economists love nothing better than when deep and liquid markets operate under conditions of perfect information,” says Jessica Irvine, National Economics Editor for News Corp Australia.

This leads us to the topic of this chapter, an introduction to the world of making decisions, processing information, and understanding behavior in markets—the world of economics. Each chapter in this book will start with a discussion about current (or sometimes past) events and revisit it at chapter’s end—to “bring home” the concepts in play.

**Note:****Introduction**

In this chapter, you will learn about:

- What Is Economics, and Why Is It Important?
- Microeconomics and Macroeconomics
- How Economists Use Theories and Models to Understand Economic Issues
- How Economies Can Be Organized: An Overview of Economic Systems

What is economics and why should you spend your time learning it? After all, there are other disciplines you could be studying, and other ways you could be spending your time. As the Bring it Home feature just mentioned, making choices is at the heart of what economists study, and your decision to take this course is as much as economic decision as anything else.

Economics is probably not what you think. It is not primarily about money or finance. It is not primarily about business. It is not mathematics. What is it then? It is both a subject area and a way of viewing the world.

## What Economics Is and Why It's Important

By the end of this section, you will be able to:

- Discuss the importance of studying economics
- Explain the relationship between production and division of labor
- Evaluate the significance of scarcity

**Economics** is the study of how humans make decisions in the face of scarcity. These can be individual decisions, family decisions, business decisions or societal decisions. If you look around carefully, you will see that scarcity is a fact of life. **Scarcity** means that human wants for goods, services and resources exceed what is available. Resources, such as labor, tools, land, and raw materials are necessary to produce the goods and services we want but they exist in limited supply. Of course, the ultimate scarce resource is time- everyone, rich or poor, has just 24 hours in the day to try to acquire the goods they want. At any point in time, there is only a finite amount of resources available.

Think about it this way: In 2015 the labor force in the United States contained over 158.6 million workers, according to the U.S. Bureau of Labor Statistics. Similarly, the total area of the United States is 3,794,101 square miles. These are large numbers for such crucial resources, however, they are limited. Because these resources are limited, so are the numbers of goods and services we produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

Scarcity of Resources



Homeless people are a stark reminder that scarcity of resources is real. (Credit: “daveynin”/Flickr Creative Commons)

If you still do not believe that scarcity is a problem, consider the following: Does everyone need food to eat? Does everyone need a decent place to live? Does everyone have access to healthcare? In every country in the world, there are people who are hungry, homeless (for example, those who call park benches their beds, as shown in [\[link\]](#)), and in need of healthcare, just to focus on a few critical goods and services. Why is this the case? It is because of scarcity. Let’s delve into the concept of scarcity a little deeper, because it is crucial to understanding economics.

## **The Problem of Scarcity**

Think about all the things you consume: food, shelter, clothing, transportation, healthcare, and entertainment. How do you acquire those items? You do not produce them yourself. You buy them. How do you afford the things you buy? You work for pay. Or if you do not, someone else does on your behalf. Yet most of us never have enough to buy all the things we want. This is because of scarcity. So how do we solve it?

**Note:**

Visit this [website](#) to read about how the United States is dealing with scarcity in resources.



Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Towns must choose whether to put more of the budget into police and fire protection or into the school system. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn't enough money in the budget to do everything. So why do we not each just produce all of the things we consume? The simple answer is most of us do not know how, but that is not the main reason. (When you study economics, you will discover that the obvious choice is not always the right answer—or at least the complete answer. Studying economics teaches you to think in a different of way.) Think back to pioneer days, when individuals knew how to do so much more than we do today, from building their homes, to growing their crops, to hunting for food, to repairing their equipment. Most of us do not know how to do all—or any—of those things. It is not because we could not learn. Rather, we do not have to. The reason why is something called *the division and specialization of labor*, a production innovation first put forth by Adam Smith, [\[link\]](#), in his book, *The Wealth of Nations*.

Adam Smith



Adam Smith  
introduced the idea  
of dividing labor  
into discrete tasks.

(Credit:  
Wikimedia  
Commons)

## The Division of and Specialization of Labor

The formal study of economics began when Adam Smith (1723–1790) published his famous book *The Wealth of Nations* in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the **division of labor**, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being done by the same person.

To illustrate the division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and

packaging pins for sale, to name just a few. Smith counted 18 distinct tasks that were often done by different people—all for a pin, believe it or not!

Modern businesses divide tasks as well. Even a relatively simple business like a restaurant divides up the task of serving meals into a range of jobs like top chef, sous chefs, less-skilled kitchen help, servers to wait on the tables, a greeter at the door, janitors to clean up, and a business manager to handle paychecks and bills—not to mention the economic connections a restaurant has with suppliers of food, furniture, kitchen equipment, and the building where it is located. A complex business like a large manufacturing factory, such as the shoe factory shown in [\[link\]](#), or a hospital can have hundreds of job classifications.

### Division of Labor



Workers on an assembly line are an example of the divisions of labor. (Credit: Nina Hale/Flickr Creative Commons)

## Why the Division of Labor Increases Production

When the tasks involved with producing a good or service are divided and subdivided, workers and businesses can produce a greater quantity of output. In his observations of pin factories, Smith observed that one worker

alone might make 20 pins in a day, but that a small business of 10 workers (some of whom would need to do two or three of the 18 tasks involved with pin-making), could make 48,000 pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process where they have an advantage. (In later chapters, we will develop this idea by discussing comparative advantage.) People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are in turn shaped by interests and talents. Only those with medical degrees qualify to become doctors, for instance. For some goods, specialization will be affected by geography—it is easier to be a wheat farmer in North Dakota than in Florida, but easier to run a tourist hotel in Florida than in North Dakota. If you live in or near a big city, it is easier to attract enough customers to operate a successful dry cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more productive than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better.

A similar pattern often operates within businesses. In many cases, a business that focuses on one or a few products (sometimes called its “core competency”) is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that for many goods, as the level of production increases, the average cost of producing each individual unit declines. For

example, if a factory produces only 100 cars per year, each car will be quite expensive to make on average. However, if a factory produces 50,000 cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will be lower. The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of their own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

## **Trade and Markets**

Specialization only makes sense, though, if workers can use the pay they receive for doing their jobs to purchase the other goods and services that they need. In short, specialization requires trade.

You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music and listen. You do not have to know anything about artificial fibers or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. You do not need to know anything about internal combustion engines to operate a car—you just get in and drive. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

## **Why Study Economics?**

Now that we have gotten an overview on what economics studies, let's quickly discuss why you are right to study it. Economics is not primarily a collection of facts to be memorized, though there are plenty of important concepts to be learned. Instead, economics is better thought of as a collection of questions to be answered or puzzles to be worked out. Most important, economics provides the tools to work out those puzzles. If you

have yet to be bitten by the economics “bug,” there are other reasons why you should study economics.

- Virtually every major problem facing the world today, from global warming, to world poverty, to the conflicts in Syria, Afghanistan, and Somalia, has an economic dimension. If you are going to be part of solving those problems, you need to be able to understand them. Economics is crucial.
- It is hard to overstate the importance of economics to good citizenship. You need to be able to vote intelligently on budgets, regulations, and laws in general. When the U.S. government came close to a standstill at the end of 2012 due to the “fiscal cliff,” what were the issues involved? Did you know?
- A basic understanding of economics makes you a well-rounded thinker. When you read articles about economic issues, you will understand and be able to evaluate the writer’s argument. When you hear classmates, co-workers, or political candidates talking about economics, you will be able to distinguish between common sense and nonsense. You will find new ways of thinking about current events and about personal and business decisions, as well as current events and politics.

The study of economics does not dictate the answers, but it can illuminate the different choices.

## **Key Concepts and Summary**

Economics seeks to solve the problem of scarcity, which is when human wants for goods and services exceed the available supply. A modern economy displays a division of labor, in which people earn income by specializing in what they produce and then use that income to purchase the products they need or want. The division of labor allows individuals and firms to specialize and to produce more for several reasons: a) It allows the agents to focus on areas of advantage due to natural factors and skill levels; b) It encourages the agents to learn and invent; c) It allows agents to take advantage of economies of scale. Division and specialization of labor only work when individuals can purchase what they do not produce in markets.

Learning about economics helps you understand the major problems facing the world today, prepares you to be a good citizen, and helps you become a well-rounded thinker.

## Self-Check Questions

### Exercise:

**Problem:** What is scarcity? Can you think of two causes of scarcity?

---

#### **Solution:**

Scarcity means human wants for goods and services exceed the available supply. Supply is limited because resources are limited. Demand, however, is virtually unlimited. Whatever the supply, it seems human nature to want more.

### Exercise:

#### **Problem:**

Residents of the town of Smithfield like to consume hams, but each ham requires 10 people to produce it and takes a month. If the town has a total of 100 people, what is the maximum amount of ham the residents can consume in a month?

---

#### **Solution:**

$100 \text{ people} / 10 \text{ people per ham} = \text{a maximum of 10 hams per month}$  if all residents produce ham. Since consumption is limited by production, the maximum number of hams residents could consume per month is 10.

### Exercise:

**Problem:**

A consultant works for \$200 per hour. She likes to eat vegetables, but is not very good at growing them. Why does it make more economic sense for her to spend her time at the consulting job and shop for her vegetables?

---

**Solution:**

She is very productive at her consulting job, but not very productive growing vegetables. Time spent consulting would produce far more income than it what she could save growing her vegetables using the same amount of time. So on purely economic grounds, it makes more sense for her to maximize her income by applying her labor to what she does best (i.e. specialization of labor).

**Exercise:****Problem:**

A computer systems engineer could paint his house, but it makes more sense for him to hire a painter to do it. Explain why.

---

**Solution:**

The engineer is better at computer science than at painting. Thus, his time is better spent working for pay at his job and paying a painter to paint his house. Of course, this assumes he does not paint his house for fun!

**Review Questions****Exercise:****Problem:**

Give the three reasons that explain why the division of labor increases an economy's level of production.

### **Exercise:**

**Problem:** What are three reasons to study economics?

## **Critical Thinking Questions**

### **Exercise:**

#### **Problem:**

Suppose you have a team of two workers: one is a baker and one is a chef. Explain why the kitchen can produce more meals in a given period of time if each worker specializes in what they do best than if each worker tries to do everything from appetizer to dessert.

### **Exercise:**

**Problem:** Why would division of labor without trade not work?

### **Exercise:**

#### **Problem:**

Can you think of any examples of *free* goods, that is, goods or services that are not scarce?

## **References**

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "The Employment Situation—February 2015." Accessed March 27, 2015.  
<http://www.bls.gov/news.release/pdf/empstat.pdf>.

Williamson, Lisa. "US Labor Market in 2012." *Bureau of Labor Statistics*. Accessed December 1, 2013.  
<http://www.bls.gov/opub/mlr/2013/03/art1full.pdf>.

## **Glossary**

division of labor

the way in which the work required to produce a good or service is divided into tasks performed by different workers

economics

the study of how humans make choices under conditions of scarcity

economies of scale

when the average cost of producing each individual unit declines as total output increases

scarcity

when human wants for goods and services exceed the available supply

specialization

when workers or firms focus on particular tasks for which they are well-suited within the overall production process

## Microeconomics and Macroeconomics

By the end of this section, you will be able to:

- Describe microeconomics
- Describe macroeconomics
- Contrast monetary policy and fiscal policy

Economics is concerned with the well-being of *all* people, including those with jobs and those without jobs, as well as those with high incomes and those with low incomes. Economics acknowledges that production of useful goods and services can create problems of environmental pollution. It explores the question of how investing in education helps to develop workers' skills. It probes questions like how to tell when big businesses or big labor unions are operating in a way that benefits society as a whole and when they are operating in a way that benefits their owners or members at the expense of others. It looks at how government spending, taxes, and regulations affect decisions about production and consumption.

It should be clear by now that economics covers a lot of ground. That ground can be divided into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses; **Macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth of production, the number of unemployed people, the inflationary increase in prices, government deficits, and levels of exports and imports. Microeconomics and macroeconomics are not separate subjects, but rather complementary perspectives on the overall subject of the economy.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the entire ecosystem of the lake from top to bottom; what eats what, how the system stays in a rough balance, and what environmental stresses affect this balance. Both approaches are useful, and both examine the same lake, but the viewpoints are different. In a

similar way, both microeconomics and macroeconomics study the same economy, but each has a different viewpoint.

Whether you are looking at lakes or economics, the micro and the macro insights should blend with each other. In studying a lake, the micro insights about particular plants and animals help to understand the overall food chain, while the macro insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by whether the macroeconomy is healthy; for example, firms will be more likely to hire workers if the overall economy is growing. In turn, the performance of the macroeconomy ultimately depends on the microeconomic decisions made by individual households and businesses.

## **Microeconomics**

What determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

What determines the products, and how many of each, a firm will produce and sell? What determines what prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomic part of this book, we will learn about the theory of consumer behavior and the theory of the firm.

## **Macroeconomics**

What determines the level of economic activity in a society? In other words, what determines how many goods and services a nation actually produces? What determines how many jobs are available in an economy? What

determines a nation's standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or to lay workers off? Finally, what causes the economy to grow over the long term?

An economy's macroeconomic health can be defined by a number of goals: growth in the standard of living, low unemployment, and low inflation, to name the most important. How can macroeconomic policy be used to pursue these goals? **Monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets, is conducted by a nation's central bank. For the United States, this is the Federal Reserve. **Fiscal policy**, which involves government spending and taxes, is determined by a nation's legislative body. For the United States, this is the Congress and the executive branch, which originates the federal budget. These are the main tools the government has to work with. Americans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that will be explored in the macroeconomic chapters of this book.

## Key Concepts and Summary

Microeconomics and macroeconomics are two different perspectives on the economy. The microeconomic perspective focuses on parts of the economy: individuals, firms, and industries. The macroeconomic perspective looks at the economy as a whole, focusing on goals like growth in the standard of living, unemployment, and inflation. Macroeconomics has two types of policies for pursuing these goals: monetary policy and fiscal policy.

## Self-Check Questions

### Exercise:

#### Problem:

What would be another example of a “system” in the real world that could serve as a metaphor for micro and macroeconomics?

---

**Solution:**

There are many physical systems that would work, for example, the study of planets (micro) in the solar system (macro), or solar systems (micro) in the galaxy (macro).

**Review Questions****Exercise:****Problem:**

What is the difference between microeconomics and macroeconomics?

**Exercise:**

**Problem:** What are examples of individual economic agents?

**Exercise:**

**Problem:** What are the three main goals of macroeconomics?

**Critical Thinking Questions****Exercise:****Problem:**

A balanced federal budget and a balance of trade are considered secondary goals of macroeconomics, while growth in the standard of living (for example) is considered a primary goal. Why do you think that is so?

**Exercise:**

**Problem:**

Macroeconomics is an aggregate of what happens at the microeconomic level. Would it be possible for what happens at the macro level to differ from how economic agents would react to some stimulus at the micro level? *Hint:* Think about the behavior of crowds.

**Glossary**

fiscal policy

economic policies that involve government spending and taxes

macroeconomics

the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance.

microeconomics

the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and business firms

monetary policy

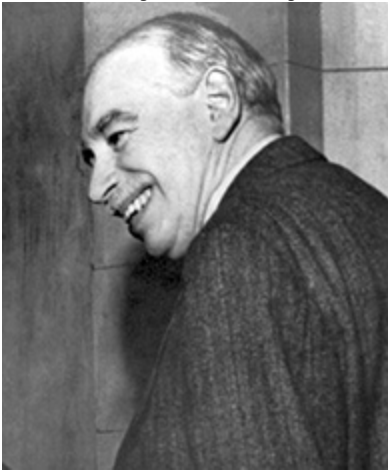
policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

## How Economists Use Theories and Models to Understand Economic Issues

By the end of this section, you will be able to:

- Interpret a circular flow diagram
- Explain the importance of economic theories and models
- Describe goods and services markets and labor markets

### John Maynard Keynes



One of the most  
influential  
economists in  
modern times was  
John Maynard  
Keynes. (Credit:  
Wikimedia  
Commons)

John Maynard Keynes (1883–1946), one of the greatest economists of the twentieth century, pointed out that economics is not just a subject area but also a way of thinking. Keynes, shown in [\[link\]](#), famously wrote in the introduction to a fellow economist's book: “[Economics] is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which

helps its possessor to draw correct conclusions.” In other words, economics teaches you how to think, not what to think.

**Note:**

Watch this [video](#) about John Maynard Keynes and his influence on economics.



Economists see the world through a different lens than anthropologists, biologists, classicists, or practitioners of any other discipline. They analyze issues and problems with economic theories that are based on particular assumptions about human behavior, that are different than the assumptions an anthropologist or psychologist might use. A **theory** is a simplified representation of how two or more variables interact with each other. The purpose of a theory is to take a complex, real-world issue and simplify it down to its essentials. If done well, this enables the analyst to understand the issue and any problems around it. A good theory is simple enough to be understood, while complex enough to capture the key features of the object or situation being studied.

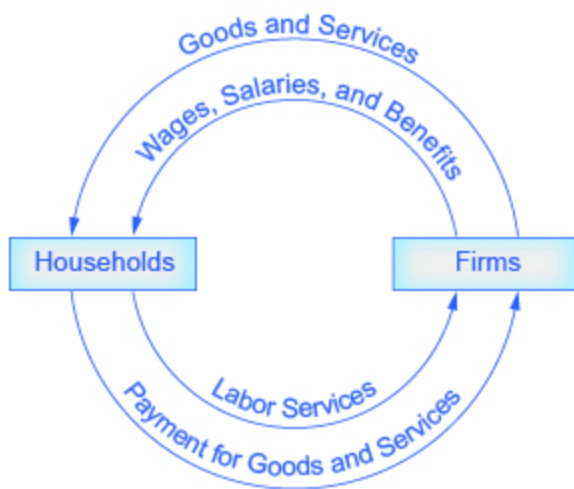
Sometimes economists use the term **model** instead of theory. Strictly speaking, a theory is a more abstract representation, while a model is more applied or empirical representation. Models are used to test theories, but for this course we will use the terms interchangeably.

For example, an architect who is planning a major office building will often build a physical model that sits on a tabletop to show how the entire city

block will look after the new building is constructed. Companies often build models of their new products, which are more rough and unfinished than the final product will be, but can still demonstrate how the new product will work.

A good model to start with in economics is the **circular flow diagram**, which is shown in [\[link\]](#). It pictures the economy as consisting of two groups—households and firms—that interact in two markets: the **goods and services market** in which firms sell and households buy and the **labor market** in which households sell labor to business firms or other employees.

### The Circular Flow Diagram



The circular flow diagram shows how households and firms interact in the goods and services market, and in the labor market. The direction of the arrows shows that in the goods and services market, households receive goods and services and pay firms for them. In the labor market, households provide labor and receive payment from firms through wages, salaries, and benefits.

Of course, in the real world, there are many different markets for goods and services and markets for many different types of labor. The circular flow diagram simplifies this to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in return for revenues. This is shown in the outer circle, and represents the two sides of the product market (for example, the market for goods and services) in which households demand and firms supply. Households sell their labor as workers to firms in return for wages, salaries and benefits. This is shown in the inner circle and represents the two sides of the labor market in which households supply and firms demand.

This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the product and labor markets work in the economy. We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, and interactions with the rest of the globe (imports and exports).

Economists carry a set of theories in their heads like a carpenter carries around a toolkit. When they see an economic issue or problem, they go through the theories they know to see if they can find one that fits. Then they use the theory to derive insights about the issue or problem. In economics, theories are expressed as diagrams, graphs, or even as mathematical equations. (Do not worry. In this course, we will mostly use graphs.) Economists do not figure out the answer to the problem first and then draw the graph to illustrate. Rather, they use the graph of the theory to help them figure out the answer. Although at the introductory level, you can sometimes figure out the right answer without applying a model, if you keep studying economics, before too long you will run into issues and problems that you will need to graph to solve. Both micro and macroeconomics are explained in terms of theories and models. The most well-known theories are probably those of supply and demand, but you will learn a number of others.

## **Key Concepts and Summary**

Economists analyze problems differently than do other disciplinary experts. The main tools economists use are economic theories or models. A theory is not an illustration of the answer to a problem. Rather, a theory is a tool for determining the answer.

## **Self-Check Questions**

### **Exercise:**

#### **Problem:**

Suppose we extend the circular flow model to add imports and exports. Copy the circular flow diagram onto a sheet of paper and then add a foreign country as a third agent. Draw a rough sketch of the flows of imports, exports, and the payments for each on your diagram.

---

#### **Solution:**

Draw a box outside the original circular flow to represent the foreign country. Draw an arrow from the foreign country to firms, to represent imports. Draw an arrow in the reverse direction representing payments for imports. Draw an arrow from firms to the foreign country to represent exports. Draw an arrow in the reverse direction to represent payments for imports.

### **Exercise:**

#### **Problem:**

What is an example of a problem in the world today, not mentioned in the chapter, that has an economic dimension?

---

#### **Solution:**

There are many such problems. Consider the AIDS epidemic. Why are so few AIDS patients in Africa and Southeast Asia treated with the same drugs that are effective in the United States and Europe? It is because neither those patients nor the countries in which they live have the resources to purchase the same drugs.

## Review Questions

### Exercise:

**Problem:** How did John Maynard Keynes define economics?

### Exercise:

#### **Problem:**

Are households primarily buyers or sellers in the goods and services market? In the labor market?

### Exercise:

#### **Problem:**

Are firms primarily buyers or sellers in the goods and services market? In the labor market?

## Critical Thinking Questions

### Exercise:

#### **Problem:**

Why is it unfair or meaningless to criticize a theory as “unrealistic?”

### Exercise:

#### **Problem:**

Suppose, as an economist, you are asked to analyze an issue unlike anything you have ever done before. Also, suppose you do not have a specific model for analyzing that issue. What should you do? *Hint:* What would a carpenter do in a similar situation?

## Glossary

circular flow diagram

a diagram that views the economy as consisting of households and firms interacting in a goods and services market and a labor market

goods and services market

a market in which firms are sellers of what they produce and households are buyers

labor market

the market in which households sell their labor as workers to business firms or other employers

model

see theory

theory

a representation of an object or situation that is simplified while including enough of the key features to help us understand the object or situation

## How Economies Can Be Organized: An Overview of Economic Systems

By the end of this section, you will be able to:

- Contrast traditional economies, command economies, and market economies
- Explain gross domestic product (GDP)
- Assess the importance and effects of globalization

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society provides is the same as the amount it needs and wants? Who insures that the right number of employees work in the electronics industry? Who insures that televisions are produced in the best way possible? How does it all get done?

There are at least three ways societies have found to organize an economy. The first is the **traditional economy**, which is the oldest economic system and can be found in parts of Asia, Africa, and South America. Traditional economies organize their economic affairs the way they have always done (i.e., tradition). Occupations stay in the family. Most families are farmers who grow the crops they have always grown using traditional methods. What you produce is what you get to consume. Because things are driven by tradition, there is little economic progress or development.

**A Command Economy**



Ancient Egypt was an example of a command economy. (Credit: Jay Bergesen/Flickr Creative Commons)

Command economies are very different. In a **command economy**, economic effort is devoted to goals passed down from a ruler or ruling class. Ancient Egypt was a good example: a large part of economic life was devoted to building pyramids, like those shown in [\[link\]](#), for the pharaohs. Medieval manor life is another example: the lord provided the land for growing crops and protection in the event of war. In return, vassals provided labor and soldiers to do the lord's bidding. In the last century, communism emphasized command economies.

In a command economy, the government decides what goods and services will be produced and what prices will be charged for them. The government decides what methods of production will be used and how much workers will be paid. Many necessities like healthcare and education are provided for free. Currently, Cuba and North Korea have command economies.

**A Market Economy**



Nothing says  
“market” more  
than The New  
York Stock  
Exchange. (Credit:  
Erik Drost/Flickr  
Creative  
Commons)

Although command economies have a very centralized structure for economic decisions, market economies have a very decentralized structure. A **market** is an institution that brings together buyers and sellers of goods or services, who may be either individuals or businesses. The New York Stock Exchange, shown in [\[link\]](#), is a prime example of market in which buyers and sellers are brought together. In a **market economy**, decision-making is decentralized. Market economies are based on **private enterprise**: the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals. Businesses supply goods and services based on demand. (In a command economy, by contrast, resources and businesses are owned by the government.) What goods and services are supplied depends on what is demanded. A person’s income is based on his or her ability to convert resources (especially labor) into something that society values. The more society values the person’s output, the higher the income (think Lady Gaga

or LeBron James). In this scenario, economic decisions are determined by market forces, not governments.

Most economies in the real world are mixed; they combine elements of command and market (and even traditional) systems. The U.S. economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America, while primarily market-oriented, have a greater degree of government involvement in economic decisions than does the U.S. economy. China and Russia, while they are closer to having a market-oriented system now than several decades ago, remain closer to the command economy end of the spectrum. A rich resource of information about countries and their economies can be found on the Heritage Foundation's website, as the following Clear It Up feature discusses.

**Note:**

**What countries are considered economically free?**

Who is in control of economic decisions? Are people free to do what they want and to work where they want? Are businesses free to produce when they want and what they choose, and to hire and fire as they wish? Are banks free to choose who will receive loans? Or does the government control these kinds of choices? Each year, researchers at the Heritage Foundation and the *Wall Street Journal* look at 50 different categories of economic freedom for countries around the world. They give each nation a score based on the extent of economic freedom in each category.

The 2015 Heritage Foundation's Index of Economic Freedom report ranked 178 countries around the world: some examples of the most free and the least free countries are listed in [\[link\]](#). Several countries were not ranked because of extreme instability that made judgments about economic freedom impossible. These countries include Afghanistan, Iraq, Syria, and Somalia.

The assigned rankings are inevitably based on estimates, yet even these rough measures can be useful for discerning trends. In 2015, 101 of the 178 included countries shifted toward greater economic freedom, although 77 of the countries shifted toward less economic freedom. In recent

decades, the overall trend has been a *higher level of economic freedom around the world*.

<b>Most Economic Freedom</b>	<b>Least Economic Freedom</b>
1. Hong Kong	167. Timor-Leste
2. Singapore	168. Democratic Republic of Congo
3. New Zealand	169. Argentina
4. Australia	170. Republic of Congo
5. Switzerland	171. Iran
6. Canada	172. Turkmenistan
7. Chile	173. Equatorial Guinea
8. Estonia	174. Eritrea
9. Ireland	175. Zimbabwe
10. Mauritius	176. Venezuela
11. Denmark	177. Cuba
12. United States	178. North Korea

Economic Freedoms, 2015(Source: The Heritage Foundation, 2015 Index of Economic Freedom, Country Rankings, <http://www.heritage.org/index/ranking>)

---

## Regulations: The Rules of the Game

Markets and government regulations are always entangled. There is no such thing as an absolutely free market. Regulations always define the “rules of the game” in the economy. Economies that are primarily market-oriented have fewer regulations—ideally just enough to maintain an even playing field for participants. At a minimum, these laws govern matters like safeguarding private property against theft, protecting people from violence, enforcing legal contracts, preventing fraud, and collecting taxes. Conversely, even the most command-oriented economies operate using markets. How else would buying and selling occur? But the decisions of what will be produced and what prices will be charged are heavily regulated. Heavily regulated economies often have **underground economies**, which are markets where the buyers and sellers make transactions without the government’s approval.

The question of how to organize economic institutions is typically not a black-or-white choice between all market or all government, but instead involves a balancing act over the appropriate combination of market freedom and government rules.

Globalization



Cargo ships are one mode of transportation

for shipping goods in the global economy.  
(Credit: Raul Valdez/Flickr Creative Commons)

## The Rise of Globalization

Recent decades have seen a trend toward **globalization**, which is the expanding cultural, political, and economic connections between people around the world. One measure of this is the increased buying and selling of goods, services, and assets across national borders—in other words, international trade and financial capital flows.

Globalization has occurred for a number of reasons. Improvements in shipping, as illustrated by the container ship shown in [\[link\]](#), and air cargo have driven down transportation costs. Innovations in computing and telecommunications have made it easier and cheaper to manage long-distance economic connections of production and sales. Many valuable products and services in the modern economy can take the form of information—for example: computer software; financial advice; travel planning; music, books and movies; and blueprints for designing a building. These products and many others can be transported over telephones and computer networks at ever-lower costs. Finally, international agreements and treaties between countries have encouraged greater trade.

[\[link\]](#) presents one measure of globalization. It shows the percentage of domestic economic production that was exported for a selection of countries from 2010 to 2013, according to an entity known as The World Bank. **Exports** are the goods and services that are produced domestically and sold abroad. **Imports** are the goods and services that are produced abroad and then sold domestically. The size of total production in an economy is measured by the **gross domestic product (GDP)**. Thus, the ratio of exports divided by GDP measures what share of a country's total economic production is sold in other countries.

Country	2010	2011	2012	2013
<b>Higher Income Countries</b>				
United States	12.4	13.6	13.6	13.5
Belgium	76.2	81.4	82.2	82.8
Canada	29.1	30.7	30.0	30.1
France	26.0	27.8	28.1	28.3
<b>Middle Income Countries</b>				
Brazil	10.9	11.9	12.6	12.6
Mexico	29.9	31.2	32.6	31.7
South Korea	49.4	55.7	56.3	53.9
<b>Lower Income Countries</b>				
Chad	36.8	38.9	36.9	32.2
China	29.4	28.5	27.3	26.4
India	22.0	23.9	24.0	24.8
Nigeria	25.3	31.3	31.4	18.0

The Extent of Globalization (exports/GDP)(Source: <http://databank.worldbank.org/data/>)

In recent decades, the export/GDP ratio has generally risen, both worldwide and for the U.S. economy. Interestingly, the share of U.S. exports in proportion to the U.S. economy is well below the global average, in part because large economies like the United States can contain more of the

division of labor inside their national borders. However, smaller economies like Belgium, Korea, and Canada need to trade across their borders with other countries to take full advantage of division of labor, specialization, and economies of scale. In this sense, the enormous U.S. economy is less affected by globalization than most other countries.

[\[link\]](#) also shows that many medium and low income countries around the world, like Mexico and China, have also experienced a surge of globalization in recent decades. If an astronaut in orbit could put on special glasses that make all economic transactions visible as brightly colored lines and look down at Earth, the astronaut would see the planet covered with connections.

So, hopefully, you now have an idea of what economics is about. Before you move to any other chapter of study, be sure to read the very important appendix to this chapter called [The Use of Mathematics in Principles of Economics](#). It is essential that you learn more about how to read and use models in economics.

**Note:****Decisions ... Decisions in the Social Media Age**

The world we live in today provides nearly instant access to a wealth of information. Consider that as recently as the late 1970s, the Farmer's Almanac, along with the Weather Bureau of the U.S. Department of Agriculture, were the primary sources American farmers used to determine when to plant and harvest their crops. Today, farmers are more likely to access, online, weather forecasts from the National Oceanic and Atmospheric Administration or watch the Weather Channel. After all, knowing the upcoming forecast could drive when to harvest crops. Consequently, knowing the upcoming weather could change the amount of crop harvested.

Some relatively new information forums, such as Facebook, are rapidly changing how information is distributed; hence, influencing decision making. In 2014, the Pew Research Center reported that 71% of online adults use Facebook. Facebook post topics range from the National Basketball Association, to celebrity singers and performers, to farmers.

Information helps us make decisions. Decisions as simple as what to wear today to how many reporters should be sent to cover a crash. Each of these decisions is an economic decision. After all, resources are scarce. If ten reporters are sent to cover an accident, they are not available to cover other stories or complete other tasks. Information provides the knowledge needed to make the best possible decisions on how to utilize scarce resources. Welcome to the world of economics!

## Key Concepts and Summary

Societies can be organized as traditional, command, or market-oriented economies. Most societies are a mix. The last few decades have seen globalization evolve as a result of growth in commercial and financial networks that cross national borders, making businesses and workers from different economies increasingly interdependent.

## Self-Check Questions

### Exercise:

#### Problem:

The chapter defines *private enterprise* as a characteristic of market-oriented economies. What would *public enterprise* be? *Hint*: It is a characteristic of command economies.

---

#### Solution:

Public enterprise means the factors of production (resources and businesses) are owned and operated by the government.

### Exercise:

#### Problem:

Why might Belgium, France, Italy, and Sweden have a higher export to GDP ratio than the United States?

---

---

**Solution:**

The United States is a large country economically speaking, so it has less need to trade internationally than the other countries mentioned. (This is the same reason that France and Italy have lower ratios than Belgium or Sweden.) One additional reason is that each of the other countries is a member of the European Union, where trade between members occurs without barriers to trade, like tariffs and quotas.

**Review Questions****Exercise:****Problem:**

What are the three ways that societies can organize themselves economically?

**Exercise:****Problem:**

What is globalization? How do you think it might have affected the economy over the past decade?

**Critical Thinking Questions****Exercise:****Problem:**

Why do you think that most modern countries' economies are a mix of command and market types?

**Exercise:**

## **Problem:**

Can you think of ways that globalization has helped you economically? Can you think of ways that it has not?

## **References**

The Heritage Foundation. 2015. "2015 Index of Economic Freedom." Accessed March 11, 2015. <http://www.heritage.org/index/ranking>.

Garling, Caleb. "S.F. plane crash: Reporting, emotions on social media," *The San Francisco Chronicle*. July 7, 2013. <http://www.sfgate.com/news/article/S-F-plane-crash-Reporting-emotions-on-social-4651639.php>.

Irvine, Jessica. "Social Networking Sites are Factories of Modern Ideas." *The Sydney Morning Herald*. November 25, 2011. <http://www.smh.com.au/federal-politics/society-and-culture/social-networking-sites-are-factories-of-modern-ideas-20111124-1nwy3.html#ixzz2YZhPYeME>.

Pew Research Center. 2015. "Social Networking Fact Sheet." Accessed March 11, 2015. <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/>.

The World Bank Group. 2015. "World Data Bank." Accessed March 30, 2014. <http://databank.worldbank.org/data/>.

## **Glossary**

command economy

an economy where economic decisions are passed down from government authority and where resources are owned by the government

exports

products (goods and services) made domestically and sold abroad

globalization

the trend in which buying and selling in markets have increasingly crossed national borders

gross domestic product (GDP)

measure of the size of total production in an economy

imports

products (goods and services) made abroad and then sold domestically

market

interaction between potential buyers and sellers; a combination of demand and supply

market economy

an economy where economic decisions are decentralized, resources are owned by private individuals, and businesses supply goods and services based on demand

private enterprise

system where the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals

traditional economy

typically an agricultural economy where things are done the same as they have always been done

underground economy

a market where the buyers and sellers make transactions in violation of one or more government regulations

Introduction to Choice in a World of Scarcity  
class="introduction"  
Choices and Tradeoffs

In general, the  
higher the degree,  
the higher the salary.  
So why aren't more  
people pursuing  
higher degrees? The  
short answer:  
choices and  
tradeoffs. (Credit:  
modification of  
work by "Jim, the  
Photographer"/Flick  
r Creative  
Commons)



**Note:****Choices ... To What Degree?**

In 2015, the median income for workers who hold master's degrees varies from males to females. The average of the two is \$2,951 weekly. Multiply this average by 52 weeks, and you get an average salary of \$153,452.

Compare that to the median weekly earnings for a full-time worker over 25 with no higher than a bachelor's degree: \$1,224 weekly and \$63,648 a year. What about those with no higher than a high school diploma in 2015? They earn just \$664 weekly and \$34,528 over 12 months. In other words, says the Bureau of Labor Statistics (BLS), earning a bachelor's degree boosted salaries 54% over what you would have earned if you had stopped your education after high school. A master's degree yields a salary almost double that of a high school diploma.

Given these statistics, we might expect a lot of people to choose to go to college and at least earn a bachelor's degree. Assuming that people want to improve their material well-being, it seems like they would make those choices that give them the greatest opportunity to consume goods and services. As it turns out, the analysis is not nearly as simple as this. In fact, in 2014, the BLS reported that while almost 88% of the population in the United States had a high school diploma, only 33.6% of 25–65 year olds had bachelor's degrees, and only 7.4% of 25–65 year olds in 2014 had earned a master's.

This brings us to the subject of this chapter: why people make the choices they make and how economists go about explaining those choices.

**Note:****Introduction to Choice in a World of Scarcity**

In this chapter, you will learn about:

- How Individuals Make Choices Based on Their Budget Constraint
- The Production Possibilities Frontier and Social Choices
- Confronting Objections to the Economic Approach

You will learn quickly when you examine the relationship between economics and scarcity that choices involve tradeoffs. Every choice has a cost.

In 1968, the Rolling Stones recorded “You Can’t Always Get What You Want.” Economists chuckled, because they had been singing a similar tune for decades. English economist Lionel Robbins (1898–1984), in his *Essay on the Nature and Significance of Economic Science* in 1932, described not always getting what you want in this way:

"The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. ... Everywhere we turn, if we choose one thing we must relinquish others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy given ends is an almost ubiquitous condition of human nature."

Because people live in a world of scarcity, they cannot have all the time, money, possessions, and experiences they wish. Neither can society.

This chapter will continue our discussion of scarcity and the economic way of thinking by first introducing three critical concepts: opportunity cost, marginal decision making, and diminishing returns. Later, it will consider whether the economic way of thinking accurately describes either how choices *are* made or how they *should* be made.

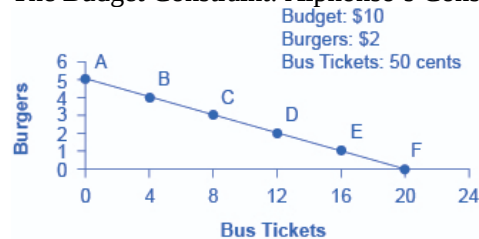
## How Individuals Make Choices Based on Their Budget Constraint

By the end of this section, you will be able to:

- Calculate and graph budget constraints
- Explain opportunity sets and opportunity costs
- Evaluate the law of diminishing marginal utility
- Explain how marginal analysis and utility influence choices

Consider the typical consumer's budget problem. Consumers have a limited amount of income to spend on the things they need and want. Suppose Alphonso has \$10 in spending money each week that he can allocate between bus tickets for getting to work and the burgers that he eats for lunch. Burgers cost \$2 each, and bus tickets are 50 cents each. [\[link\]](#) shows Alphonso's **budget constraint**, that is, the outer boundary of his **opportunity set**. The opportunity set identifies all the opportunities for spending within his budget. The budget constraint indicates all the combinations of burgers and bus tickets Alphonso can afford when he exhausts his budget, given the prices of the two goods. (There are actually many different kinds of budget constraints. You will learn more about them in the chapter on [Consumer Choices](#).)

**The Budget Constraint: Alphonso's Consumption Choice Opportunity Frontier**



Each point on the budget constraint represents a combination of burgers and bus tickets whose total cost adds up to Alphonso's budget of \$10. The slope of the budget constraint is determined by the relative price of burgers and bus tickets. All along the budget set, giving up one burger means gaining four bus tickets.

The vertical axis in the figure shows burger purchases and the horizontal axis shows bus ticket purchases. If Alphonso spends all his money on burgers, he can afford five per week. ( $\$10 \text{ per week} / \$2 \text{ per burger} = 5 \text{ burgers per week}$ .) But if he does this, he will not be able to afford any bus tickets. This choice (zero bus tickets and five burgers) is shown by point A in the figure. Alternatively, if Alphonso spends all his money on bus tickets, he can afford 20 per week. ( $\$10 \text{ per week} / \$0.50 \text{ per bus ticket} = 20 \text{ bus tickets per week}$ .) Then, however, he will not be able to afford any burgers. This alternative choice (20 bus tickets and zero burgers) is shown by point F.

If Alphonso is like most people, he will choose some combination that includes both bus tickets and burgers. That is, he will choose some combination on the budget constraint that connects points A and F. Every point on (or inside) the constraint shows a combination of burgers and bus tickets that Alphonso can afford. Any point outside the constraint is not affordable, because it would cost more money than Alphonso has in his budget.

The budget constraint clearly shows the tradeoff Alphonso faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and two burgers. What would it cost Alphonso for one more burger? It would be natural to answer \$2, but that's not the way economists think. Instead they ask, how many bus tickets would Alphonso have to give up to get one more burger, while staying within his budget? The answer is four bus tickets. That is the true cost to Alphonso of one more burger.

## The Concept of Opportunity Cost

Economists use the term **opportunity cost** to indicate what must be given up to obtain something that is desired. The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else; in short, opportunity cost is the value of the next best alternative. For Alphonso, the opportunity cost of a burger is the four bus tickets he would have to give up. He would decide whether or not to choose the burger depending on whether the value of the burger exceeds the value of the forgone alternative—in this case, bus tickets. Since people must choose, they inevitably face tradeoffs in which they have to give up things they desire to get other things they desire more.

### Note:

View this [website](#) for an example of opportunity cost—paying someone else to wait in line for you.



A fundamental principle of economics is that every choice has an opportunity cost. If you sleep through your economics class (not recommended, by the way), the opportunity cost is the learning you miss from not attending class. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us and part of human existence.

The following Work It Out feature shows a step-by-step analysis of a budget constraint calculation. Read through it to understand another important concept—slope—that is further explained in the appendix [The Use of Mathematics in Principles of Economics](#).

### Note:

#### Understanding Budget Constraints

Budget constraints are easy to understand if you apply a little math. The appendix [The Use of Mathematics in Principles of Economics](#) explains all the math you are likely to need in this book. So if math is not your strength, you might want to take a look at the appendix.

Step 1: The equation for any budget constraint is:

#### Equation:

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$

where P and Q are the price and quantity of items purchased and Budget is the amount of income one has to spend.

Step 2. Apply the budget constraint equation to the scenario. In Alphonso's case, this works out to be:

**Equation:**

$$\text{Budget} = P_1 \times Q_1 + P_2 \times Q_2$$

\$10 budget = \$2 per burger  $\times$  quantity of burgers + \$0.50 per bus ticket  $\times$  quantity of bus tickets

$$\$10 = \$2 \times Q_{\text{burgers}} + \$0.50 \times Q_{\text{bus tickets}}$$

Step 3. Using a little algebra, we can turn this into the familiar equation of a line:

**Equation:**

$$y = b + mx$$

For Alphonso, this is:

**Equation:**

$$\$10 = \$2 \times Q_{\text{burgers}} + \$0.50 \times Q_{\text{bus tickets}}$$

Step 4. Simplify the equation. Begin by multiplying both sides of the equation by 2:

**Equation:**

$$2 \times 10 = 2 \times 2 \times Q_{\text{burgers}} + 2 \times 0.5 \times Q_{\text{bus tickets}}$$

$$20 = 4 \times Q_{\text{burgers}} + 1 \times Q_{\text{bus tickets}}$$

Step 5. Subtract one bus ticket from both sides:

**Equation:**

$$20 - Q_{\text{bus tickets}} = 4 \times Q_{\text{burgers}}$$

Divide each side by 4 to yield the answer:

**Equation:**

$$5 - 0.25 \times Q_{\text{bus tickets}} = Q_{\text{burgers}}$$

or

$$Q_{\text{burgers}} = 5 - 0.25 \times Q_{\text{bus tickets}}$$

Step 6. Notice that this equation fits the budget constraint in [\[link\]](#). The vertical intercept is 5 and the slope is  $-0.25$ , just as the equation says. If you plug 20 bus tickets into the equation, you get 0 burgers. If you plug other numbers of bus tickets into the equation, you get the results shown in [\[link\]](#), which are the points on Alphonso's budget constraint.

Point	Quantity of Burgers (at \$2)	Quantity of Bus Tickets (at 50 cents)
A	5	0
B	4	4

Point	Quantity of Burgers (at \$2)	Quantity of Bus Tickets (at 50 cents)
C	3	8
D	2	12
E	1	16
F	0	20

Step 7. Notice that the slope of a budget constraint always shows the opportunity cost of the good which is on the horizontal axis. For Alphonso, the slope is  $-0.25$ , indicating that for every four bus tickets he buys, Alphonso must give up 1 burger.

There are two important observations here. First, the algebraic sign of the slope is negative, which means that the only way to get more of one good is to give up some of the other. Second, the slope is defined as the price of bus tickets (whatever is on the horizontal axis in the graph) divided by the price of burgers (whatever is on the vertical axis), in this case  $\$0.50/\$2 = 0.25$ . So if you want to determine the opportunity cost quickly, just divide the two prices.

## Identifying Opportunity Cost

In many cases, it is reasonable to refer to the opportunity cost as the price. If your cousin buys a new bicycle for \$300, then \$300 measures the amount of “other consumption” that he has given up. For practical purposes, there may be no special need to identify the specific alternative product or products that could have been bought with that \$300, but sometimes the price as measured in dollars may not accurately capture the true opportunity cost. This problem can loom especially large when costs of time are involved.

For example, consider a boss who decides that all employees will attend a two-day retreat to “build team spirit.” The out-of-pocket monetary cost of the event may involve hiring an outside consulting firm to run the retreat, as well as room and board for all participants. But an opportunity cost exists as well: during the two days of the retreat, none of the employees are doing any other work.

Attending college is another case where the opportunity cost exceeds the monetary cost. The out-of-pocket costs of attending college include tuition, books, room and board, and other expenses. But in addition, during the hours that you are attending class and studying, it is impossible to work at a paying job. Thus, college imposes both an out-of-pocket cost and an opportunity cost of lost earnings.

### Note:

What is the opportunity cost associated with increased airport security measures?

After the terrorist plane hijackings on September 11, 2001, many steps were proposed to improve air travel safety. For example, the federal government could provide armed “sky marshals” who would travel inconspicuously with the rest of the passengers. The cost of having a sky marshal on every flight would be roughly \$3 billion per year. Retrofitting all U.S. planes with reinforced cockpit doors to make it harder for terrorists to take over the plane would have a price tag of \$450 million. Buying more sophisticated security equipment for airports, like three-dimensional baggage scanners and cameras linked to face recognition software, could cost another \$2 billion.

But the single biggest cost of greater airline security does not involve spending money. It is the opportunity cost of additional waiting time at the airport. According to the United States Department of Transportation

(DOT), more than 800 million passengers took plane trips in the United States in 2012. Since the 9/11 hijackings, security screening has become more intensive, and consequently, the procedure takes longer than in the past. Say that, on average, each air passenger spends an extra 30 minutes in the airport per trip. Economists commonly place a value on time to convert an opportunity cost in time into a monetary figure. Because many air travelers are relatively high-paid business people, conservative estimates set the average price of time for air travelers at \$20 per hour. By these back-of-the-envelope calculations, the opportunity cost of delays in airports could be as much as  $800 \text{ million} \times 0.5 \text{ hours} \times \$20/\text{hour}$ , or \$8 billion per year. Clearly, the opportunity costs of waiting time can be just as important as costs that involve direct spending.

In some cases, realizing the opportunity cost can alter behavior. Imagine, for example, that you spend \$8 on lunch every day at work. You may know perfectly well that bringing a lunch from home would cost only \$3 a day, so the opportunity cost of buying lunch at the restaurant is \$5 each day (that is, the \$8 buying lunch costs minus the \$3 your lunch from home would cost). \$5 each day does not seem to be that much. However, if you project what that adds up to in a year—250 days a year  $\times$  \$5 per day equals \$1,250, the cost, perhaps, of a decent vacation. If the opportunity cost is described as “a nice vacation” instead of “\$5 a day,” you might make different choices.

## Marginal Decision-Making and Diminishing Marginal Utility

The budget constraint framework helps to emphasize that most choices in the real world are not about getting all of one thing or all of another; that is, they are not about choosing either the point at one end of the budget constraint or else the point all the way at the other end. Instead, most choices involve **marginal analysis**, which means comparing the benefits and costs of choosing a little more or a little less of a good.

People desire goods and services for the satisfaction or **utility** those goods and services provide. Utility, as we will see in the chapter on [Consumer Choices](#), is subjective but that does not make it less real. Economists typically assume that the more of some good one consumes (for example, slices of pizza), the more utility one obtains. At the same time, the utility a person receives from consuming the first unit of a good is typically more than the utility received from consuming the fifth or the tenth unit of that same good. When Alphonso chooses between burgers and bus tickets, for example, the first few bus rides that he chooses might provide him with a great deal of utility—perhaps they help him get to a job interview or a doctor’s appointment. But later bus rides might provide much less utility—they may only serve to kill time on a rainy day. Similarly, the first burger that Alphonso chooses to buy may be on a day when he missed breakfast and is ravenously hungry. However, if Alphonso has a burger every single day, the last few burgers may taste pretty boring. The general pattern that consumption of the first few units of any good tends to bring a higher level of utility to a person than consumption of later units is a common pattern. Economists refer to this pattern as the **law of diminishing marginal utility**, which means that as a person receives more of a good, the additional (or marginal) utility from each additional unit of the good declines. In other words, the first slice of pizza brings more satisfaction than the sixth.

The law of diminishing marginal utility explains why people and societies rarely make all-or-nothing choices. You would not say, “My favorite food is ice cream, so I will eat nothing but ice cream from now on.” Instead, even if you get a very high level of utility from your favorite food, if you ate it exclusively, the additional or marginal utility from those last few servings would not be very high. Similarly, most workers do not say: “I enjoy leisure, so I’ll never work.” Instead, workers recognize that even though some leisure is very nice, a combination of all leisure and no income is not so attractive. The budget constraint framework suggests that when people make choices in a world of scarcity, they will use marginal analysis and think about whether they would prefer a little more or a little less.

## Sunk Costs

In the budget constraint framework, all decisions involve what will happen next: that is, what quantities of goods will you consume, how many hours will you work, or how much will you save. These decisions do not look back to past choices. Thus, the budget constraint framework assumes that **sunk costs**, which are costs that were incurred in the past and cannot be recovered, should not affect the current decision.

Consider the case of Selena, who pays \$8 to see a movie, but after watching the film for 30 minutes, she knows that it is truly terrible. Should she stay and watch the rest of the movie because she paid for the ticket, or should she leave? The money she spent is a sunk cost, and unless the theater manager is feeling kindly, Selena will not get a refund. But staying in the movie still means paying an opportunity cost in time. Her choice is whether to spend the next 90 minutes suffering through a cinematic disaster or to do something—anything—else. The lesson of sunk costs is to forget about the money and time that is irretrievably gone and instead to focus on the marginal costs and benefits of current and future options.

For people and firms alike, dealing with sunk costs can be frustrating. It often means admitting an earlier error in judgment. Many firms, for example, find it hard to give up on a new product that is doing poorly because they spent so much money in creating and launching the product. But the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future.

## **From a Model with Two Goods to One of Many Goods**

The budget constraint diagram containing just two goods, like most models used in this book, is not realistic. After all, in a modern economy people choose from thousands of goods. However, thinking about a model with many goods is a straightforward extension of what we discussed here. Instead of drawing just one budget constraint, showing the tradeoff between two goods, you can draw multiple budget constraints, showing the possible tradeoffs between many different pairs of goods. Or in more advanced classes in economics, you would use mathematical equations that include many possible goods and services that can be purchased, together with their quantities and prices, and show how the total spending on all goods and services is limited to the overall budget available. The graph with two goods that was presented here clearly illustrates that every choice has an opportunity cost, which is the point that does carry over to the real world.

## **Key Concepts and Summary**

Economists see the real world as one of scarcity: that is, a world in which people's desires exceed what is possible. As a result, economic behavior involves tradeoffs in which individuals, firms, and society must give up something that they desire to obtain things that they desire more. Individuals face the tradeoff of what quantities of goods and services to consume. The budget constraint, which is the frontier of the opportunity set, illustrates the range of choices available. The slope of the budget constraint is determined by the relative price of the choices. Choices beyond the budget constraint are not affordable.

Opportunity cost measures cost by what is given up in exchange. Sometimes opportunity cost can be measured in money, but it is often useful to consider time as well, or to measure it in terms of the actual resources that must be given up.

Most economic decisions and tradeoffs are not all-or-nothing. Instead, they involve marginal analysis, which means they are about decisions on the margin, involving a little more or a little less. The law of diminishing marginal utility points out that as a person receives more of something—whether it is a specific good or another resource—the additional marginal gains tend to become smaller. Because sunk costs occurred in the past and cannot be recovered, they should be disregarded in making current decisions.

## **Self-Check Questions**

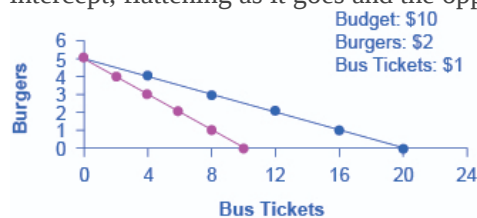
### Exercise:

#### Problem:

Suppose Alphonso's town raised the price of bus tickets to \$1 per trip (while the price of burgers stayed at \$2 and his budget remained \$10 per week.) Draw Alphonso's new budget constraint. What happens to the opportunity cost of bus tickets?

#### Solution:

The opportunity cost of bus tickets is the number of burgers that must be given up to obtain one more bus ticket. Originally, when the price of bus tickets was 50 cents per trip, this opportunity cost was  $0.50/2 = .25$  burgers. The reason for this is that at the original prices, one burger (\$2) costs the same as four bus tickets (\$0.50), so the opportunity cost of a burger is four bus tickets, and the opportunity cost of a bus ticket is .25 (the inverse of the opportunity cost of a burger). With the new, higher price of bus tickets, the opportunity cost rises to  $1/2$  or 0.50. You can see this graphically since the slope of the new budget constraint is flatter than the original one. If Alphonso spends all of his budget on burgers, the higher price of bus tickets has no impact so the horizontal intercept of the budget constraint is the same. If he spends all of his budget on bus tickets, he can now afford only half as many, so the vertical intercept is half as much. In short, the budget constraint rotates clockwise around the horizontal intercept, flattening as it goes and the opportunity cost of bus tickets increases.



### Review Questions

#### Exercise:

**Problem:** Explain why scarcity leads to tradeoffs.

#### Exercise:

#### Problem:

Explain why individuals make choices that are directly on the budget constraint, rather than inside the budget constraint or outside it.

### Critical Thinking Question

#### Exercise:

#### Problem:

Suppose Alphonso's town raises the price of bus tickets from \$0.50 to \$1 and the price of burgers rises from \$2 to \$4. Why is the opportunity cost of bus tickets unchanged? Suppose Alphonso's weekly spending money increases from \$10 to \$20. How is his budget constraint affected from all three changes? Explain.

## Problems

Use this information to answer the following 4 questions: Marie has a weekly budget of \$24, which she likes to spend on magazines and pies.

### Exercise:

#### Problem:

If the price of a magazine is \$4 each, what is the maximum number of magazines she could buy in a week?

### Exercise:

**Problem:** If the price of a pie is \$12, what is the maximum number of pies she could buy in a week?

### Exercise:

#### Problem:

Draw Marie's budget constraint with pies on the horizontal axis and magazines on the vertical axis. What is the slope of the budget constraint?

### Exercise:

**Problem:** What is Marie's opportunity cost of purchasing a pie?

## References

Bureau of Labor Statistics, U.S. Department of Labor. 2015. "Median Weekly Earnings by Educational Attainment in 2014." Accessed March 27, 2015. <http://www.bls.gov/opub/ted/2015/median-weekly-earnings-by-education-gender-race-and-ethnicity-in-2014.htm>.

Robbins, Lionel. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan. 1932.

United States Department of Transportation. "Total Passengers on U.S Airlines and Foreign Airlines U.S. Flights Increased 1.3% in 2012 from 2011." Accessed October 2013. [http://www.rita.dot.gov/bts/press\\_releases/bts016\\_13](http://www.rita.dot.gov/bts/press_releases/bts016_13)

## Glossary

budget constraint

all possible consumption combinations of goods that someone can afford, given the prices of goods, when all income is spent; the boundary of the opportunity set

law of diminishing marginal utility

as we consume more of a good or service, the utility we get from additional units of the good or service tend to become smaller than what we received from earlier units

marginal analysis

examination of decisions on the margin, meaning a little more or a little less from the status quo

opportunity cost

measures cost by what is given up in exchange; opportunity cost measures the value of the forgone alternative

opportunity set

all possible combinations of consumption that someone can afford given the prices of goods and the individual's income

sunk costs

costs that are made in the past and cannot be recovered

utility

satisfaction, usefulness, or value one obtains from consuming goods and services

## Introduction to Demand and Supply

class="introduction"

### Farmer's Market

Organic vegetables and fruits that are grown and sold within a specific geographical region should, in theory, cost less than conventional produce because the transportation costs are less.

That is not, however, usually the case. (Credit: modification of work by Natalie Maynor/Flickr r Creative Commons)

**Note:****Why Can We Not Get Enough of Organic?**

Organic food is increasingly popular, not just in the United States, but worldwide. At one time, consumers had to go to specialty stores or farmer's markets to find organic produce. Now it is available in most grocery stores. In short, organic is part of the mainstream.

Ever wonder why organic food costs more than conventional food? Why, say, does an organic Fuji apple cost \$1.99 a pound, while its conventional counterpart costs \$1.49 a pound? The same price relationship is true for just about every organic product on the market. If many organic foods are locally grown, would they not take less time to get to market and therefore be cheaper? What are the forces that keep those prices from coming down? Turns out those forces have a lot to do with this chapter's topic: demand and supply.

**Note:****Introduction to Demand and Supply**

In this chapter, you will learn about:

- Demand, Supply, and Equilibrium in Markets for Goods and Services
- Shifts in Demand and Supply for Goods and Services
- Changes in Equilibrium Price and Quantity: The Four-Step Process
- Price Ceilings and Price Floors

An auction bidder pays thousands of dollars for a dress Whitney Houston wore. A collector spends a small fortune for a few drawings by John Lennon. People usually react to purchases like these in two ways: their jaw drops because they think these are high prices to pay for such goods or they think these are rare, desirable items and the amount paid seems right.

**Note:**

Visit this [website](#) to read a list of bizarre items that have been purchased for their ties to celebrities. These examples represent an interesting facet of demand and supply.



When economists talk about prices, they are less interested in making judgments than in gaining a practical understanding of what determines prices and why prices change. Consider a price most of us contend with weekly: that of a gallon of gas. Why was the average price of gasoline in the United States \$3.71 per gallon in June 2014? Why did the price for gasoline fall sharply to \$2.07 per gallon by January 2015? To explain these price movements, economists focus on the determinants of what gasoline buyers are willing to pay and what gasoline sellers are willing to accept.

As it turns out, the price of gasoline in June of any given year is nearly always higher than the price in January of that same year; over recent decades, gasoline prices in midsummer have averaged about 10 cents per gallon more than their midwinter low. The likely reason is that people drive more in the summer, and are also willing to pay more for gas, but that does not explain how steeply gas prices fell. Other factors were at work during those six months, such as increases in supply and decreases in the demand for crude oil.

This chapter introduces the economic model of demand and supply—one of the most powerful models in all of economics. The discussion here begins by examining how demand and supply determine the price and the quantity sold in markets for goods and services, and how changes in demand and supply lead to changes in prices and quantities.

## Demand, Supply, and Equilibrium in Markets for Goods and Services

By the end of this section, you will be able to:

- Explain demand, quantity demanded, and the law of demand
- Identify a demand curve and a supply curve
- Explain supply, quantity supply, and the law of supply
- Explain equilibrium, equilibrium price, and equilibrium quantity

First let's first focus on what economists mean by demand, what they mean by supply, and then how demand and supply interact in a market.

### Demand for Goods and Services

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is based on needs and wants—a consumer may be able to differentiate between a need and a want, but from an economist's perspective they are the same thing. Demand is also based on ability to pay. If you cannot pay for it, you have no effective demand.

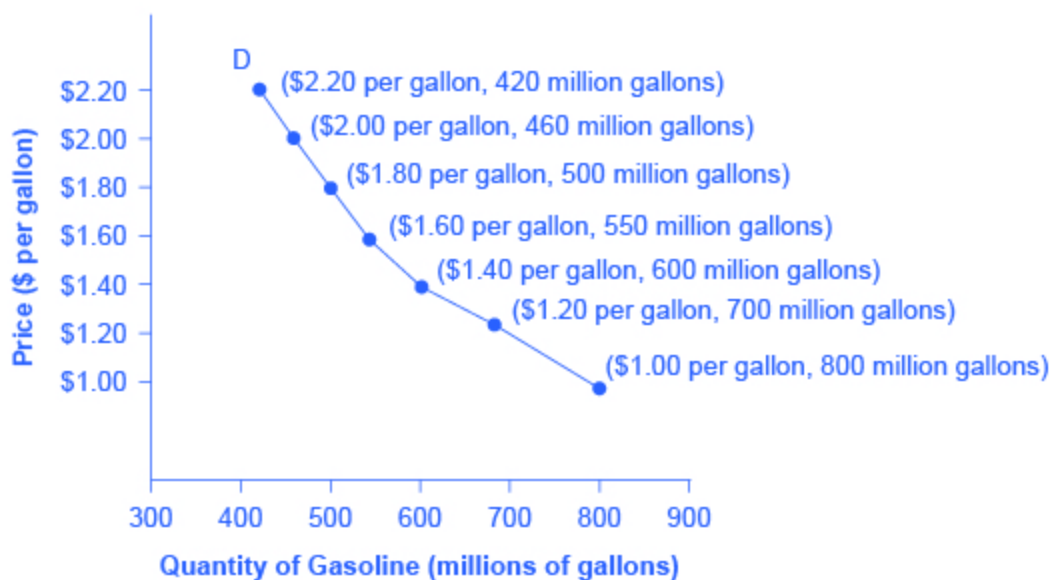
What a buyer pays for a unit of the specific good or service is called **price**. The total number of units purchased at that price is called the **quantity demanded**. A rise in price of a good or service almost always decreases the quantity demanded of that good or service. Conversely, a fall in price will increase the quantity demanded. When the price of a gallon of gasoline goes up, for example, people look for ways to reduce their consumption by combining several errands, commuting by carpool or mass transit, or taking weekend or vacation trips closer to home. Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand (to be explained in the next module) are held constant.

An example from the market for gasoline can be shown in the form of a table or a graph. A table that shows the quantity demanded at each price, such as [\[link\]](#), is called a **demand schedule**. Price in this case is measured in dollars per gallon of gasoline. The quantity demanded is measured in

millions of gallons over some time period (for example, per day or per year) and over some geographic area (like a state or a country). A **demand curve** shows the relationship between price and quantity demanded on a graph like [\[link\]](#), with quantity on the horizontal axis and the price per gallon on the vertical axis. (Note that this is an exception to the normal rule in mathematics that the independent variable (x) goes on the horizontal axis and the dependent variable (y) goes on the vertical. Economics is not math.)

The demand schedule shown by [\[link\]](#) and the demand curve shown by the graph in [\[link\]](#) are two ways of describing the same relationship between price and quantity demanded.

### A Demand Curve for Gasoline



The demand schedule shows that as price rises, quantity demanded decreases, and vice versa. These points are then graphed, and the line connecting them is the demand curve (D). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

Price (per gallon)	Quantity Demanded (millions of gallons)
\$1.00	800
\$1.20	700
\$1.40	600
\$1.60	550
\$1.80	500
\$2.00	460
\$2.20	420

### Price and Quantity Demanded of Gasoline

Demand curves will appear somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. So demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases.

Confused about these different types of demand? Read the next Clear It Up feature.

#### **Note:**

Is demand the same as quantity demanded?

In economic terminology, demand is not the same as quantity demanded. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or

one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

## Supply of Goods and Services

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a good or service. A rise in price almost always leads to an increase in the **quantity supplied** of that good or service, while a fall in price will decrease the quantity supplied. When the price of gasoline rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants where it can be refined into gasoline; build new oil refineries; purchase additional pipelines and trucks to ship the gasoline to gas stations; and open more gas stations or keep existing gas stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply assumes that all other variables that affect supply (to be explained in the next module) are held constant.

Still unsure about the different types of supply? See the following Clear It Up feature.

### **Note:**

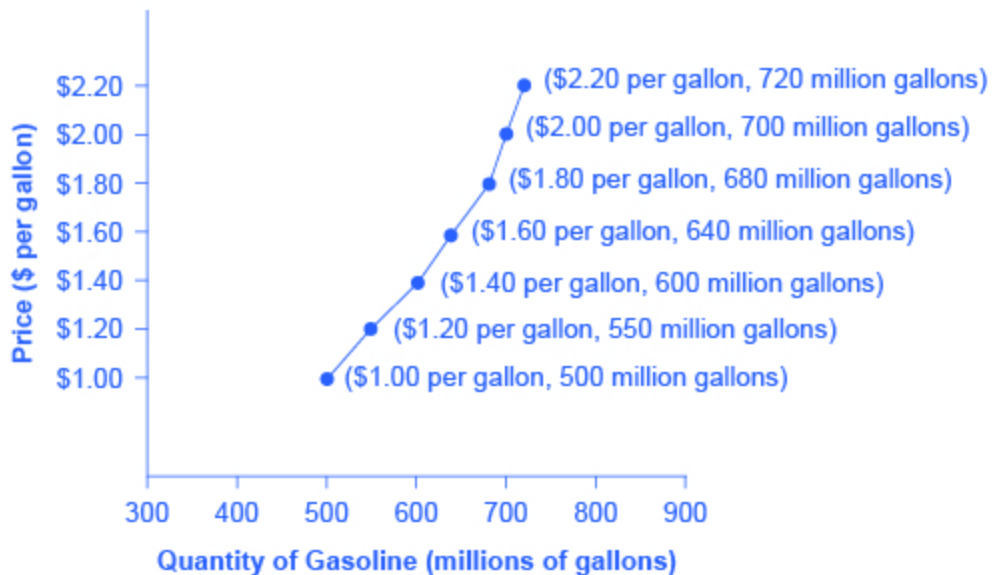
Is supply the same as quantity supplied?

In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a range of prices and the quantities supplied at those prices, a relationship that can be illustrated with a supply curve or a supply schedule. When economists refer to quantity supplied, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply

refers to the curve and quantity supplied refers to the (specific) point on the curve.

[\[link\]](#) illustrates the law of supply, again using the market for gasoline as an example. Like demand, supply can be illustrated using a table or a graph. A **supply schedule** is a table, like [\[link\]](#), that shows the quantity supplied at a range of different prices. Again, price is measured in dollars per gallon of gasoline and quantity supplied is measured in millions of gallons. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

#### A Supply Curve for Gasoline



The supply schedule is the table that shows quantity supplied of gasoline at each price. As price rises, quantity supplied also increases, and vice versa. The supply curve (S) is created by graphing the points from the supply schedule and then connecting them. The upward slope of the supply curve illustrates the law of supply—that a higher price leads to a higher quantity supplied, and vice versa.

<b>Price (per gallon)</b>	<b>Quantity Supplied (millions of gallons)</b>
\$1.00	500
\$1.20	550
\$1.40	600
\$1.60	640
\$1.80	680
\$2.00	700
\$2.20	720

### Price and Supply of Gasoline

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: they slope up from left to right and illustrate the law of supply: as the price rises, say, from \$1.00 per gallon to \$2.20 per gallon, the quantity supplied increases from 500 gallons to 720 gallons. Conversely, as the price falls, the quantity supplied decreases.

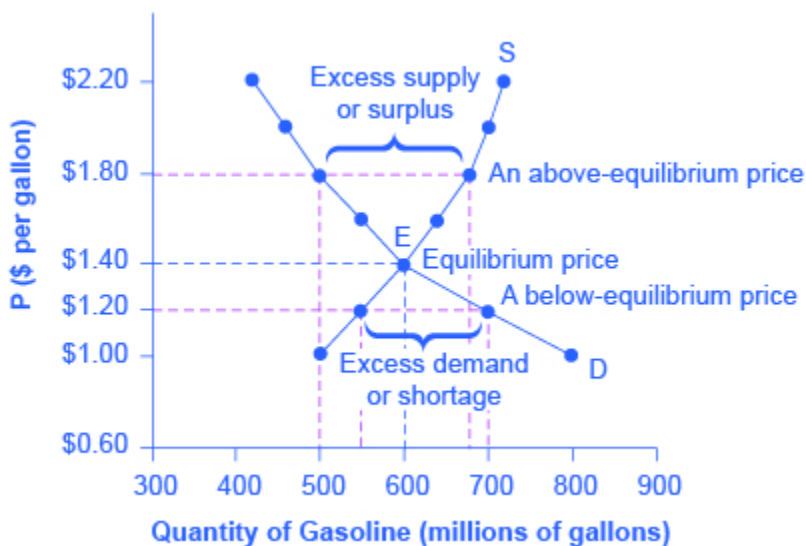
### Equilibrium—Where Demand and Supply Intersect

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph.

Together, demand and supply determine the price and the quantity that will be bought and sold in a market.

[\[link\]](#) illustrates the interaction of demand and supply in the market for gasoline. The demand curve (D) is identical to [\[link\]](#). The supply curve (S) is identical to [\[link\]](#). [\[link\]](#) contains the same information in tabular form.

### Demand and Supply for Gasoline



The demand curve (D) and the supply curve (S) intersect at the equilibrium point E, with a price of \$1.40 and a quantity of 600. The equilibrium is the only price where quantity demanded is equal to quantity supplied. At a price above equilibrium like \$1.80, quantity supplied exceeds the quantity demanded, so there is excess supply. At a price below equilibrium such as \$1.20, quantity demanded exceeds quantity supplied, so there is excess demand.

<b>Price (per gallon)</b>	<b>Quantity demanded (millions of gallons)</b>	<b>Quantity supplied (millions of gallons)</b>
\$1.00	800	500
\$1.20	700	550
<b>\$1.40</b>	<b>600</b>	<b>600</b>
\$1.60	550	640
\$1.80	500	680
\$2.00	460	700
\$2.20	420	720

### Price, Quantity Demanded, and Quantity Supplied

Remember this: When two lines on a diagram cross, this intersection usually means something. The point where the supply curve (S) and the demand curve (D) cross, designated by point E in [\[link\]](#), is called the **equilibrium**. The **equilibrium price** is the only price where the plans of consumers and the plans of producers agree—that is, where the amount of the product consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). This common quantity is called the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

In [\[link\]](#), the equilibrium price is \$1.40 per gallon of gasoline and the equilibrium quantity is 600 million gallons. If you had only the demand and supply schedules, and not the graph, you could find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal.

The word “equilibrium” means “balance.” If a market is at its equilibrium price and quantity, then it has no reason to move away from that point. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and the equilibrium quantity.

Imagine, for example, that the price of a gallon of gasoline was above the equilibrium price—that is, instead of \$1.40 per gallon, the price is \$1.80 per gallon. This above-equilibrium price is illustrated by the dashed horizontal line at the price of \$1.80 in [\[link\]](#). At this higher price, the quantity demanded drops from 600 to 500. This decline in quantity reflects how consumers react to the higher price by finding ways to use less gasoline.

Moreover, at this higher price of \$1.80, the quantity of gasoline supplied rises from the 600 to 680, as the higher price makes it more profitable for gasoline producers to expand their output. Now, consider how quantity demanded and quantity supplied are related at this above-equilibrium price. Quantity demanded has fallen to 500 gallons, while quantity supplied has risen to 680 gallons. In fact, at any above-equilibrium price, the quantity supplied exceeds the quantity demanded. We call this an **excess supply** or a **surplus**.

With a surplus, gasoline accumulates at gas stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation puts pressure on gasoline sellers. If a surplus remains unsold, those firms involved in making and selling gasoline are not receiving enough cash to pay their workers and to cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions in turn will stimulate a higher quantity demanded. So, if the price is above the equilibrium level, incentives built into the structure of demand and supply will create pressures for the price to fall toward the equilibrium.

Now suppose that the price is below its equilibrium level at \$1.20 per gallon, as the dashed horizontal line at this price in [\[link\]](#) shows. At this lower price, the quantity demanded increases from 600 to 700 as drivers take longer trips, spend more minutes warming up the car in the driveway in wintertime, stop sharing rides to work, and buy larger cars that get fewer

miles to the gallon. However, the below-equilibrium price reduces gasoline producers' incentives to produce and sell gasoline, and the quantity supplied falls from 600 to 550.

When the price is below equilibrium, there is **excess demand**, or a **shortage**—that is, at the given price the quantity demanded, which has been stimulated by the lower price, now exceeds the quantity supplied, which had been depressed by the lower price. In this situation, eager gasoline buyers mob the gas stations, only to find many stations running short of fuel. Oil companies and gas stations recognize that they have an opportunity to make higher profits by selling what gasoline they have at a higher price. As a result, the price rises toward the equilibrium level. Read [Demand, Supply, and Efficiency](#) for more discussion on the importance of the demand and supply model.

## Key Concepts and Summary

A demand schedule is a table that shows the quantity demanded at different prices in the market. A demand curve shows the relationship between quantity demanded and price in a given market on a graph. The law of demand states that a higher price typically leads to a lower quantity demanded.

A supply schedule is a table that shows the quantity supplied at different prices in the market. A supply curve shows the relationship between quantity supplied and price on a graph. The law of supply says that a higher price typically leads to a higher quantity supplied.

The equilibrium price and equilibrium quantity occur where the supply and demand curves cross. The equilibrium occurs where the quantity demanded is equal to the quantity supplied. If the price is below the equilibrium level, then the quantity demanded will exceed the quantity supplied. Excess demand or a shortage will exist. If the price is above the equilibrium level, then the quantity supplied will exceed the quantity demanded. Excess supply or a surplus will exist. In either case, economic pressures will push the price toward the equilibrium level.

## Self-Check Question

### Exercise:

#### Problem:

Review [\[link\]](#). Suppose the price of gasoline is \$1.60 per gallon. Is the quantity demanded higher or lower than at the equilibrium price of \$1.40 per gallon? And what about the quantity supplied? Is there a shortage or a surplus in the market? If so, of how much?

---

#### Solution:

Since \$1.60 per gallon is above the equilibrium price, the quantity demanded would be lower at 550 gallons and the quantity supplied would be higher at 640 gallons. (These results are due to the laws of demand and supply, respectively.) The outcome of lower  $Q_d$  and higher  $Q_s$  would be a surplus in the gasoline market of  $640 - 550 = 90$  gallons.

## Review Questions

### Exercise:

**Problem:** What determines the level of prices in a market?

### Exercise:

#### Problem:

What does a downward-sloping demand curve mean about how buyers in a market will react to a higher price?

### Exercise:

#### Problem:

Will demand curves have the same exact shape in all markets? If not, how will they differ?

**Exercise:**

**Problem:**

Will supply curves have the same shape in all markets? If not, how will they differ?

**Exercise:**

**Problem:**

What is the relationship between quantity demanded and quantity supplied at equilibrium? What is the relationship when there is a shortage? What is the relationship when there is a surplus?

**Exercise:**

**Problem:**

How can you locate the equilibrium point on a demand and supply graph?

**Exercise:**

**Problem:**

If the price is above the equilibrium level, would you predict a surplus or a shortage? If the price is below the equilibrium level, would you predict a surplus or a shortage? Why?

**Exercise:**

**Problem:**

When the price is above the equilibrium, explain how market forces move the market price to equilibrium. Do the same when the price is below the equilibrium.

**Exercise:**

**Problem:**

What is the difference between the demand and the quantity demanded of a product, say milk? Explain in words and show the difference on a graph with a demand curve for milk.

**Exercise:****Problem:**

What is the difference between the supply and the quantity supplied of a product, say milk? Explain in words and show the difference on a graph with the supply curve for milk.

**Critical Thinking Questions****Exercise:****Problem:**

Review [\[link\]](#). Suppose the government decided that, since gasoline is a necessity, its price should be legally capped at \$1.30 per gallon. What do you anticipate would be the outcome in the gasoline market?

**Exercise:****Problem:**

Explain why the following statement is false: “In the goods market, no buyer would be willing to pay more than the equilibrium price.”

**Exercise:****Problem:**

Explain why the following statement is false: “In the goods market, no seller would be willing to sell for less than the equilibrium price.”

**Problems**

## Exercise:

### Problem:

Review [\[link\]](#) again. Suppose the price of gasoline is \$1.00. Will the quantity demanded be lower or higher than at the equilibrium price of \$1.40 per gallon? Will the quantity supplied be lower or higher? Is there a shortage or a surplus in the market? If so, of how much?

## References

Costanza, Robert, and Lisa Wainger. "No Accounting For Nature: How Conventional Economics Distorts the Value of Things." *The Washington Post*. September 2, 1990.

European Commission: Agriculture and Rural Development. 2013. "Overview of the CAP Reform: 2014-2024." Accessed April 13, 2015. <http://ec.europa.eu/agriculture/cap-post-2013/>.

Radford, R. A. "The Economic Organisation of a P.O.W. Camp." *Economica*. no. 48 (1945): 189-201. <http://www.jstor.org/stable/2550133>.

## Glossary

### demand curve

a graphic representation of the relationship between price and quantity demanded of a certain good or service, with quantity on the horizontal axis and the price on the vertical axis

### demand schedule

a table that shows a range of prices for a certain good or service and the quantity demanded at each price

### demand

the relationship between price and the quantity demanded of a certain good or service

equilibrium price

the price where quantity demanded is equal to quantity supplied

equilibrium quantity

the quantity at which quantity demanded and quantity supplied are equal for a certain price level

equilibrium

the situation where quantity demanded is equal to the quantity supplied; the combination of price and quantity where there is no economic pressure from surpluses or shortages that would cause price or quantity to change

excess demand

at the existing price, the quantity demanded exceeds the quantity supplied; also called a shortage

excess supply

at the existing price, quantity supplied exceeds the quantity demanded; also called a surplus

law of demand

the common relationship that a higher price leads to a lower quantity demanded of a certain good or service and a lower price leads to a higher quantity demanded, while all other variables are held constant

law of supply

the common relationship that a higher price leads to a greater quantity supplied and a lower price leads to a lower quantity supplied, while all other variables are held constant

price

what a buyer pays for a unit of the specific good or service

quantity demanded

the total number of units of a good or service consumers are willing to purchase at a given price

quantity supplied

the total number of units of a good or service producers are willing to sell at a given price

shortage

at the existing price, the quantity demanded exceeds the quantity supplied; also called excess demand

supply curve

a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule

a table that shows a range of prices for a good or service and the quantity supplied at each price

supply

the relationship between price and the quantity supplied of a certain good or service

surplus

at the existing price, quantity supplied exceeds the quantity demanded; also called excess supply

## Shifts in Demand and Supply for Goods and Services

By the end of this section, you will be able to:

- Identify factors that affect demand
- Graph demand curves and demand shifts
- Identify factors that affect supply
- Graph supply curves and supply shifts

The previous module explored how price affects the quantity demanded and the quantity supplied. The result was the demand curve and the supply curve. Price, however, is not the only thing that influences demand. Nor is it the only thing that influences supply. For example, how is demand for vegetarian food affected if, say, health concerns cause more consumers to avoid eating meat? Or how is the supply of diamonds affected if diamond producers discover several new diamond mines? What are the major factors, in addition to the price, that influence demand or supply?

### **Note:**

Visit this [website](#) to read a brief note on how marketing strategies can influence supply and demand of products.



## What Factors Affect Demand?

We defined demand as the amount of some product a consumer is willing and able to purchase at each price. That suggests at least two factors in addition to price that affect demand. Willingness to purchase suggests a

desire, based on what economists call tastes and preferences. If you neither need nor want something, you will not buy it. Ability to purchase suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have more income. Prices of related goods can affect demand also. If you need a new car, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance, and the less for diapers and baby formula.

These factors matter both for demand by an individual and demand by the market as a whole. Exactly how do these various factors affect demand, and how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

## **The *Ceteris Paribus* Assumption**

A demand curve or a supply curve is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption **ceteris paribus**, a Latin phrase meaning “other things being equal.” Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. A demand curve or a supply curve is a relationship between two, and only two, variables when all other variables are kept constant. If all else is not held equal, then the laws of supply and demand will not necessarily hold, as the following Clear It Up feature shows.

### **Note:**

When does *ceteris paribus* apply?

*Ceteris paribus* is typically applied when we look at how changes in price affect demand or supply, but *ceteris paribus* can be applied more generally. In the real world, demand and supply depend on more factors than just

price. For example, a consumer's demand depends on income and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, assuming the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy falls for two reasons: first because of the higher price and second because of the lower income.

## How Does Income Affect Demand?

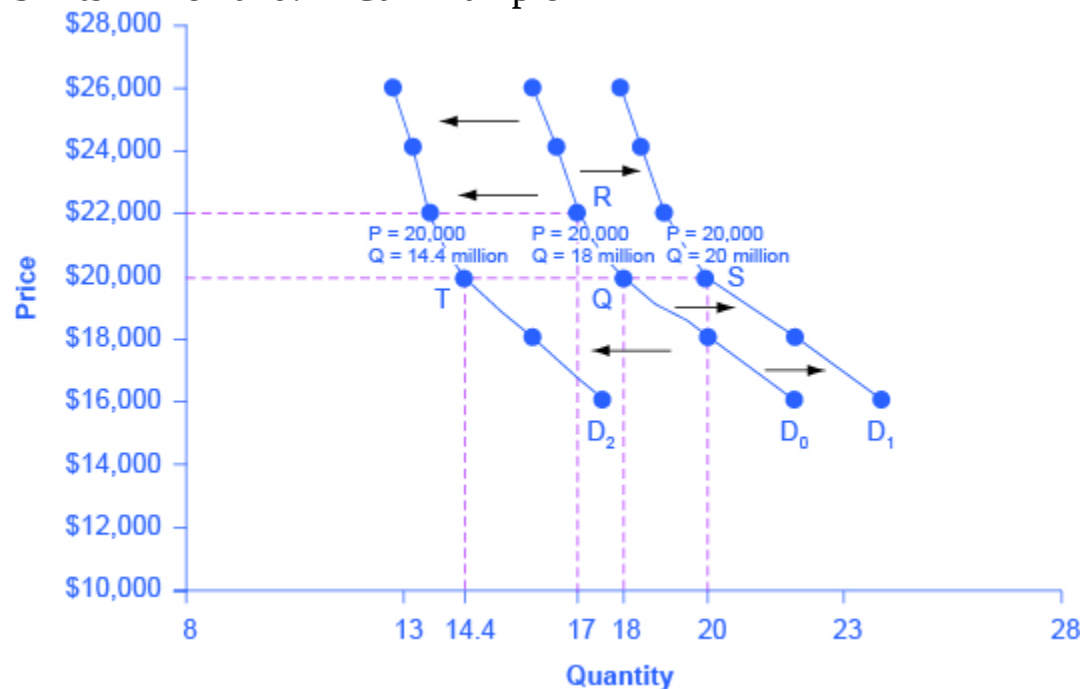
Let's use income as an example of how factors other than price affect demand. [\[link\]](#) shows the initial demand for automobiles as  $D_0$ . At point Q, for example, if the price is \$20,000 per car, the quantity of cars demanded is 18 million.  $D_0$  also shows how the quantity of cars demanded would change as a result of a higher or lower price. For example, if the price of a car rose to \$22,000, the quantity demanded would decrease to 17 million, at point R.

The original demand curve  $D_0$ , like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making cars more affordable. How will this affect demand? How can we show this graphically?

Return to [\[link\]](#). The price of cars is still \$20,000, but with higher incomes, the quantity demanded has now increased to 20 million cars, shown at point S. As a result of the higher income levels, the demand curve shifts to the right to the new demand curve  $D_1$ , indicating an increase in demand. [\[link\]](#)

shows clearly that this increased demand would occur at every price, not just the original one.

### Shifts in Demand: A Car Example



Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from  $D_0$  to  $D_1$ . Decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from  $D_0$  to  $D_2$ .

Price	Decrease to $D_2$	Original Quantity Demanded $D_0$	Increase to $D_1$
\$16,000	17.6 million	22.0 million	24.0 million

<b>Price</b>	<b>Decrease to <math>D_2</math></b>	<b>Original Quantity Demanded <math>D_0</math></b>	<b>Increase to <math>D_1</math></b>
\$18,000	16.0 million	20.0 million	22.0 million
\$20,000	14.4 million	18.0 million	20.0 million
\$22,000	13.6 million	17.0 million	19.0 million
\$24,000	13.2 million	16.5 million	18.5 million
\$26,000	12.8 million	16.0 million	18.0 million

### Price and Demand Shifts: A Car Example

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of cars demanded at every given price, and the original demand curve  $D_0$  would shift left to  $D_2$ . The shift from  $D_0$  to  $D_2$  represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of \$20,000 means 18 million cars sold along the original demand curve, but only 14.4 million sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures an pattern for the market as a whole.

In the previous section, we argued that higher income causes greater demand at every price. This is true for most goods and services. For some—

luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially pronounced. A product whose demand rises when income rises, and vice versa, is called a **normal good**. A few exceptions to this pattern do exist. As incomes rise, many people will buy fewer generic brand groceries and more name brand groceries. They are less likely to buy used cars and more likely to buy new cars. They will be less likely to rent an apartment and more likely to own a home, and so on. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

## Other Factors That Shift Demand Curves

Income is not the only factor that causes a shift in demand. Other things that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let's look at these factors.

### Changing Tastes or Preferences

From 1980 to 2014, the per-person consumption of chicken by Americans rose from 48 pounds per year to 85 pounds per year, and consumption of beef fell from 77 pounds per year to 54 pounds per year, according to the U.S. Department of Agriculture (USDA). Changes like these are largely due to movements in taste, which change the quantity of a good demanded at every price: that is, they shift the demand curve for that good, rightward for chicken and leftward for beef.

### Changes in the Composition of the Population

The proportion of elderly citizens in the United States population is rising. It rose from 9.8% in 1970 to 12.6% in 2000, and will be projected (by the U.S. Census Bureau) 20% of the population by 2030. A society with relatively more children, like the United States in the 1960s, will have greater demand for goods and services like tricycles and day care facilities.

A society with relatively more elderly persons, as the United States is projected to have by 2030, has a higher demand for nursing homes and hearing aids. Similarly, changes in the size of the population can affect the demand for housing and many other goods. Each of these changes in demand will be shown as a shift in the demand curve.

The demand for a product can also be affected by changes in the prices of related goods such as substitutes or complements. A **substitute** is a good or service that can be used in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which can be shown graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning that the goods are often used together, because consumption of one good tends to enhance consumption of the other. Examples include breakfast cereal and milk; notebooks and pens or pencils, golf balls and golf clubs; gasoline and sport utility vehicles; and the five-way combination of bacon, lettuce, tomato, mayonnaise, and bread. If the price of golf clubs rises, since the quantity demanded of golf clubs falls (because of the law of demand), demand for a complement good like golf balls decreases, too. Similarly, a higher price for skis would shift the demand curve for a complement good like ski resort trips to the left, while a lower price for a complement has the reverse effect.

### **Changes in Expectations about Future Prices or Other Factors that Affect Demand**

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that a hurricane is coming, they may rush to the store to buy flashlight batteries and bottled water. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock

up on coffee now. These changes in demand are shown as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than price) causes a different quantity to be demanded at every price. The following Work It Out feature shows how this happens.

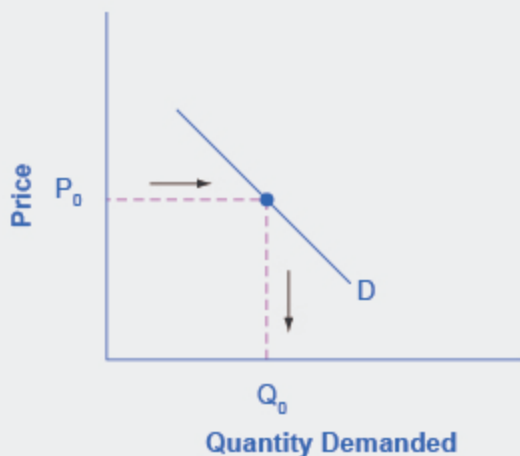
**Note:**

**Shift in Demand**

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Following is an example of a shift in demand due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza. Pick a price (like  $P_0$ ). Identify the corresponding  $Q_0$ . An example is shown in [\[link\]](#).

**Demand Curve**

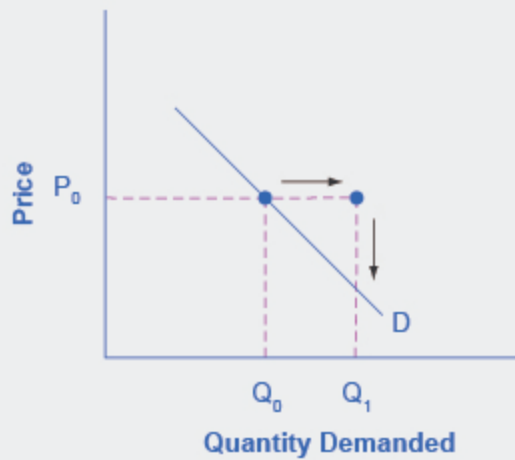


The demand curve can be used to identify how much consumers would buy at any given price.

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new  $Q_1$ . Draw a dotted vertical line

down to the horizontal axis and label the new  $Q_1$ . An example is provided in [\[link\]](#).

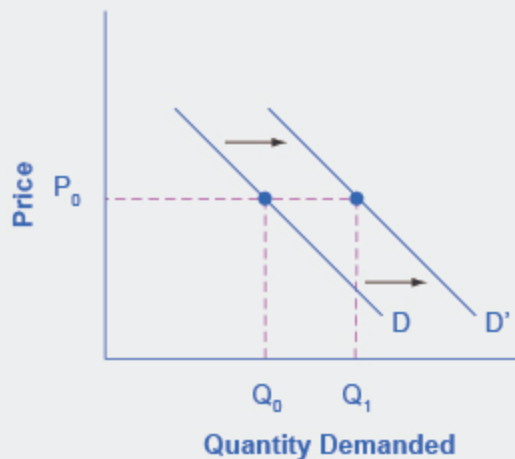
### Demand Curve with Income Increase



With an increase in income, consumers will purchase larger quantities, pushing demand to the right.

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes an upward (or rightward) shift in the demand curve, so that at any price the quantities demanded will be higher, as shown in [\[link\]](#).

### Demand Curve Shifted Right



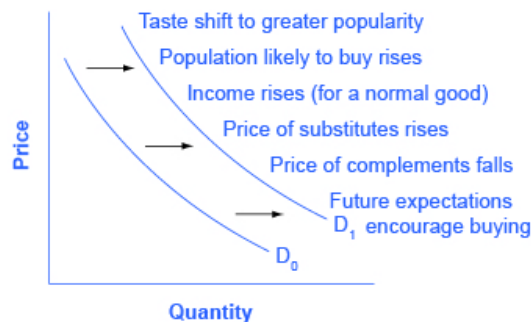
With an increase in income, consumers will

purchase larger quantities, pushing demand to the right, and causing the demand curve to shift right.

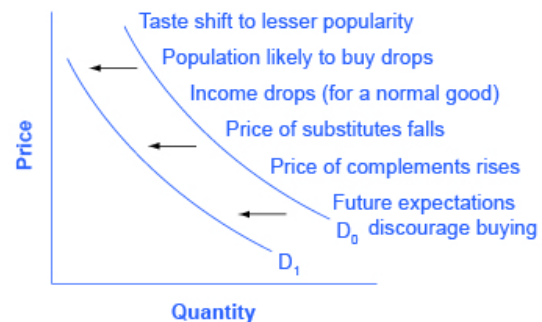
## Summing Up Factors That Change Demand

Six factors that can shift demand curves are summarized in [\[link\]](#). The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

### Factors That Shift Demand Curves



(a) Factors that increase demand



(b) Factors that decrease demand

- (a) A list of factors that can cause an increase in demand from  $D_0$  to  $D_1$ . (b) The same factors, if their direction is reversed, can cause a decrease in demand from  $D_0$  to  $D_1$ .

When a demand curve shifts, it will then intersect with a given supply curve at a different equilibrium price and quantity. We are, however, getting ahead of our story. Before discussing how changes in demand can affect

equilibrium price and quantity, we first need to discuss shifts in supply curves.

## How Production Costs Affect Supply

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus* so that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as a shift in demand is represented by a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

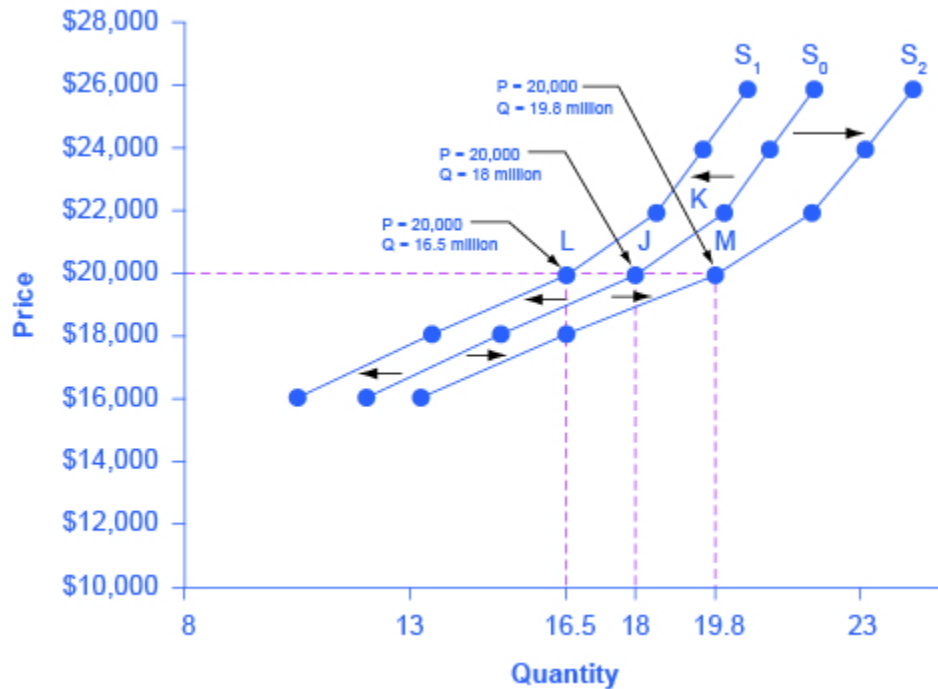
In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. Goods and services are produced using combinations of labor, materials, and machinery, or what we call **inputs** or **factors of production**. If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm's profits go up. When a firm's profits increase, it is more motivated to produce output, since the more it produces the more profit it will earn. So, when costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. This can be shown by the supply curve shifting to the right.

Take, for example, a messenger company that delivers packages around a city. The company may find that buying gasoline is one of its main costs. If the price of gasoline falls, then the company will find it can deliver messages more cheaply than before. Since lower costs correspond to higher profits, the messenger company may now supply more of its services at any given price. For example, given the lower gasoline prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply for cars, shown by curve  $S_0$  in [\[link\]](#). Point J indicates that if the price is \$20,000, the quantity supplied will be 18 million cars. If the price rises to \$22,000 per car, *ceteris paribus*, the quantity supplied will rise to 20 million cars, as point K on the  $S_0$  curve shows. The same information can be shown in table form, as in [\[link\]](#).

### Shifts in Supply: A Car Example



Decreased supply means that at every given price, the quantity supplied is lower, so that the supply curve shifts to the left, from  $S_0$  to  $S_1$ . Increased supply means that at every given price, the quantity supplied is higher, so that the supply curve shifts to the right, from  $S_0$  to  $S_2$ .

<b>Price</b>	<b>Decrease to <math>S_1</math></b>	<b>Original Quantity Supplied <math>S_0</math></b>	<b>Increase to <math>S_2</math></b>
\$16,000	10.5 million	12.0 million	13.2 million
\$18,000	13.5 million	15.0 million	16.5 million
\$20,000	16.5 million	18.0 million	19.8 million
\$22,000	18.5 million	20.0 million	22.0 million
\$24,000	19.5 million	21.0 million	23.1 million
\$26,000	20.5 million	22.0 million	24.2 million

### Price and Shifts in Supply: A Car Example

Now, imagine that the price of steel, an important ingredient in manufacturing cars, rises, so that producing a car has become more expensive. At any given price for selling cars, car manufacturers will react by supplying a lower quantity. This can be shown graphically as a leftward shift of supply, from  $S_0$  to  $S_1$ , which indicates that at any given price, the quantity supplied decreases. In this example, at a price of \$20,000, the quantity supplied decreases from 18 million on the original supply curve ( $S_0$ ) to 16.5 million on the supply curve  $S_1$ , which is labeled as point L.

Conversely, if the price of steel decreases, producing a car becomes less expensive. At any given price for selling cars, car manufacturers can now expect to earn higher profits, so they will supply a higher quantity. The shift of supply to the right, from  $S_0$  to  $S_2$ , means that at all prices, the quantity supplied has increased. In this example, at a price of \$20,000, the quantity

supplied increases from 18 million on the original supply curve ( $S_0$ ) to 19.8 million on the supply curve  $S_2$ , which is labeled M.

## **Other Factors That Affect Supply**

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

The cost of production for many agricultural products will be affected by changes in natural conditions. For example, in 2014 the Manchurian Plain in Northeastern China, which produces most of the country's wheat, corn, and soybeans, experienced its most severe drought in 50 years. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity will be supplied; conversely, especially good weather would shift the supply curve to the right.

When a firm discovers a new technology that allows the firm to produce at a lower cost, the supply curve will shift to the right, as well. For instance, in the 1960s a major scientific effort nicknamed the Green Revolution focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world was grown with these Green Revolution seeds—and the harvest was twice as high per acre. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

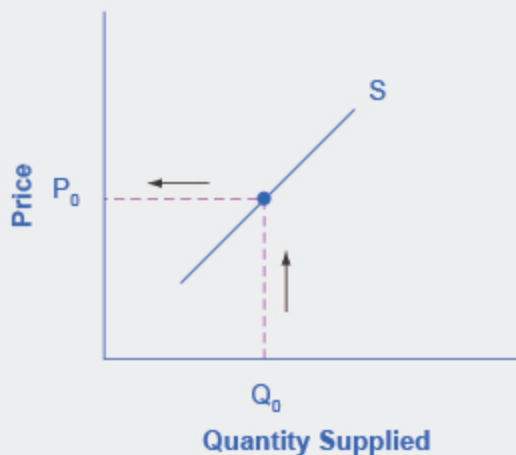
Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the U.S. government imposes a tax on alcoholic beverages that collects about \$8 billion per year from producers. Taxes are treated as costs by businesses. Higher costs decrease supply for the reasons discussed above. Other examples of policy that can affect cost are the wide array of government regulations that require firms to spend money to provide a cleaner environment or a safer workplace; complying with regulations increases costs.

A government subsidy, on the other hand, is the opposite of a tax. A subsidy occurs when the government pays a firm directly or reduces the firm's taxes if the firm carries out certain actions. From the firm's perspective, taxes or regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price. Government subsidies reduce the cost of production and increase supply at every given price, shifting supply to the right. The following Work It Out feature shows how this shift happens.

**Note:****Shift in Supply**

We know that a supply curve shows the minimum price a firm will accept to produce a given quantity of output. What happens to the supply curve when the cost of production goes up? Following is an example of a shift in supply due to a production cost increase.

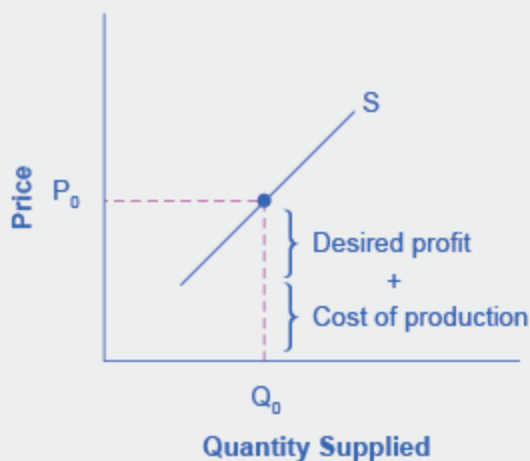
Step 1. Draw a graph of a supply curve for pizza. Pick a quantity (like  $Q_0$ ). If you draw a vertical line up from  $Q_0$  to the supply curve, you will see the price the firm chooses. An example is shown in [\[link\]](#).

**Supply Curve**

The supply curve can be used to show the minimum price a firm will accept to produce a given quantity of output.

Step 2. Why did the firm choose that price and not some other? One way to think about this is that the price is composed of two parts. The first part is the cost of producing pizzas at the margin; in this case, the cost of producing the pizza, including cost of ingredients (dough, sauce, cheese, pepperoni, and so on), the cost of the pizza oven, the rent on the shop, and the wages of the workers. The second part is the firm's desired profit, which is determined, among other factors, by the profit margins in that particular business. If you add these two parts together, you get the price the firm wishes to charge. The quantity  $Q_0$  and associated price  $P_0$  give you one point on the firm's supply curve, as shown in [\[link\]](#).

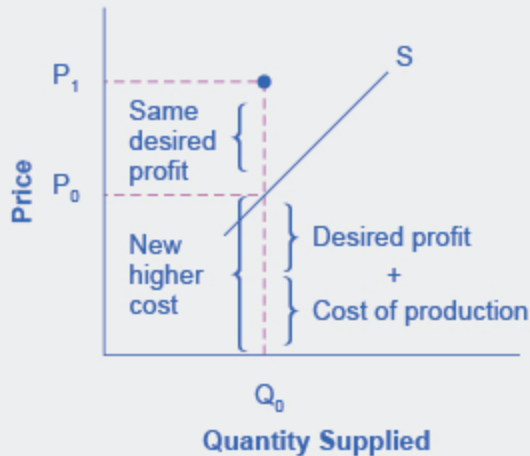
### Setting Prices



The cost of production and the desired profit equal the price a firm will set for a product.

Step 3. Now, suppose that the cost of production goes up. Perhaps cheese has become more expensive by \$0.75 per pizza. If that is true, the firm will want to raise its price by the amount of the increase in cost (\$0.75). Draw this point on the supply curve directly above the initial point on the curve, but \$0.75 higher, as shown in [\[link\]](#).

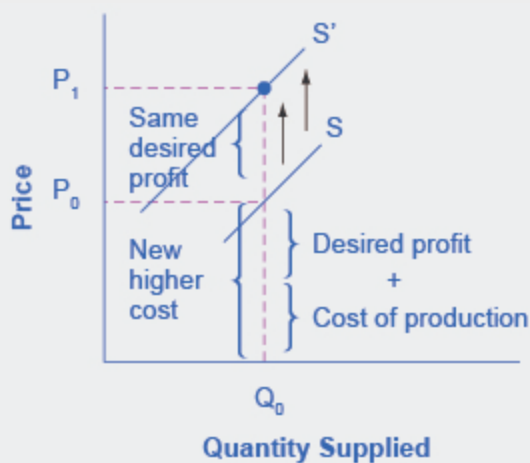
### Increasing Costs Leads to Increasing Price



Because the cost of production and the desired profit equal the price a firm will set for a product, if the cost of production increases, the price for the product will also need to increase.

Step 4. Shift the supply curve through this point. You will see that an increase in cost causes an upward (or a leftward) shift of the supply curve so that at any price, the quantities supplied will be smaller, as shown in [\[link\]](#).

### Supply Curve Shifts



When the cost of production increases, the supply curve shifts upwardly to a new price

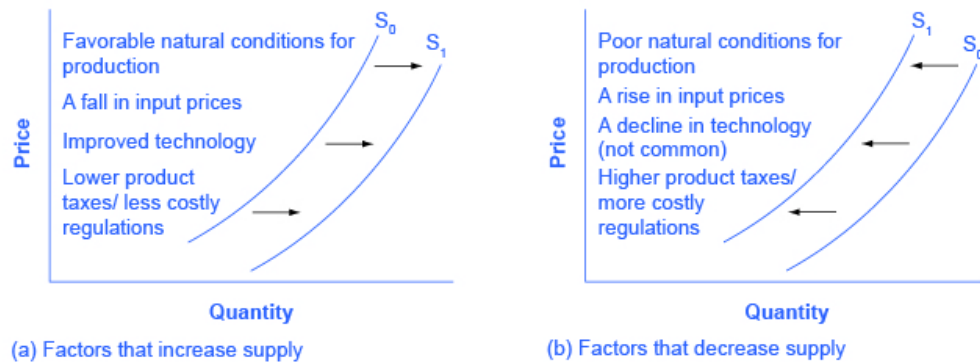
level.

## Summing Up Factors That Change Supply

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

[\[link\]](#) summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

### Factors That Shift Supply Curves



- (a) A list of factors that can cause an increase in supply from  $S_0$  to  $S_1$ .
- (b) The same factors, if their direction is reversed, can cause a decrease in supply from  $S_0$  to  $S_1$ .

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that economics is about only four

topics: demand, supply, price, and quantity. However, demand and supply are really “umbrella” concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. Factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

## Key Concepts and Summary

Economists often use the *ceteris paribus* or “other things being equal” assumption: while examining the economic impact of one event, all other factors remain unchanged for the purpose of the analysis. Factors that can shift the demand curve for goods and services, causing a different quantity to be demanded at any given price, include changes in tastes, population, income, prices of substitute or complement goods, and expectations about future conditions and prices. Factors that can shift the supply curve for goods and services, causing a different quantity to be supplied at any given price, include input prices, natural conditions, changes in technology, and government taxes, regulations, or subsidies.

## Self-Check Questions

### Exercise:

**Problem:** Why do economists use the *ceteris paribus* assumption?

---

### Solution:

To make it easier to analyze complex problems. *Ceteris paribus* allows you to look at the effect of one factor at a time on what it is you are trying to analyze. When you have analyzed all the factors individually, you add the results together to get the final answer.

### Exercise:

**Problem:**

In an analysis of the market for paint, an economist discovers the facts listed below. State whether each of these changes will affect supply or demand, and in what direction.

- a. There have recently been some important cost-saving inventions in the technology for making paint.
  - b. Paint is lasting longer, so that property owners need not repaint as often.
  - c. Because of severe hailstorms, many people need to repaint now.
  - d. The hailstorms damaged several factories that make paint, forcing them to close down for several months.
- 

**Solution:**

- a. An improvement in technology that reduces the cost of production will cause an increase in supply. Alternatively, you can think of this as a reduction in price necessary for firms to supply any quantity. Either way, this can be shown as a rightward (or downward) shift in the supply curve.
- b. An improvement in product quality is treated as an increase in tastes or preferences, meaning consumers demand more paint at any price level, so demand increases or shifts to the right. If this seems counterintuitive, note that demand in the future for the longer-lasting paint will fall, since consumers are essentially shifting demand from the future to the present.
- c. An increase in need causes an increase in demand or a rightward shift in the demand curve.
- d. Factory damage means that firms are unable to supply as much in the present. Technically, this is an increase in the cost of production. Either way you look at it, the supply curve shifts to the left.

**Exercise:**

**Problem:**

Many changes are affecting the market for oil. Predict how each of the following events will affect the equilibrium price and quantity in the market for oil. In each case, state how the event will affect the supply and demand diagram. Create a sketch of the diagram if necessary.

- a. Cars are becoming more fuel efficient, and therefore get more miles to the gallon.
- b. The winter is exceptionally cold.
- c. A major discovery of new oil is made off the coast of Norway.
- d. The economies of some major oil-using nations, like Japan, slow down.
- e. A war in the Middle East disrupts oil-pumping schedules.
- f. Landlords install additional insulation in buildings.
- g. The price of solar energy falls dramatically.
- h. Chemical companies invent a new, popular kind of plastic made from oil.

---

**Solution:**

- a. More fuel-efficient cars means there is less need for gasoline. This causes a leftward shift in the demand for gasoline and thus oil. Since the demand curve is shifting down the supply curve, the equilibrium price and quantity both fall.
- b. Cold weather increases the need for heating oil. This causes a rightward shift in the demand for heating oil and thus oil. Since the demand curve is shifting up the supply curve, the equilibrium price and quantity both rise.
- c. A discovery of new oil will make oil more abundant. This can be shown as a rightward shift in the supply curve, which will cause a decrease in the equilibrium price along with an increase in the equilibrium quantity. (The supply curve shifts down the demand curve so price and quantity follow the law of demand. If price goes down, then the quantity goes up.)

- d. When an economy slows down, it produces less output and demands less input, including energy, which is used in the production of virtually everything. A decrease in demand for energy will be reflected as a decrease in the demand for oil, or a leftward shift in demand for oil. Since the demand curve is shifting down the supply curve, both the equilibrium price and quantity of oil will fall.
- e. Disruption of oil pumping will reduce the supply of oil. This leftward shift in the supply curve will show a movement up the demand curve, resulting in an increase in the equilibrium price of oil and a decrease in the equilibrium quantity.
- f. Increased insulation will decrease the demand for heating. This leftward shift in the demand for oil causes a movement down the supply curve, resulting in a decrease in the equilibrium price and quantity of oil.
- g. Solar energy is a substitute for oil-based energy. So if solar energy becomes cheaper, the demand for oil will decrease as consumers switch from oil to solar. The decrease in demand for oil will be shown as a leftward shift in the demand curve. As the demand curve shifts down the supply curve, both equilibrium price and quantity for oil will fall.
- h. A new, popular kind of plastic will increase the demand for oil. The increase in demand will be shown as a rightward shift in demand, raising the equilibrium price and quantity of oil.

## Review Questions

### Exercise:

#### Problem:

When analyzing a market, how do economists deal with the problem that many factors that affect the market are changing at the same time?

### Exercise:

**Problem:**

Name some factors that can cause a shift in the demand curve in markets for goods and services.

**Exercise:****Problem:**

Name some factors that can cause a shift in the supply curve in markets for goods and services.

**Critical Thinking Questions****Exercise:****Problem:**

Consider the demand for hamburgers. If the price of a substitute good (for example, hot dogs) increases and the price of a complement good (for example, hamburger buns) increases, can you tell for sure what will happen to the demand for hamburgers? Why or why not? Illustrate your answer with a graph.

**Exercise:****Problem:**

How do you suppose the demographics of an aging population of “Baby Boomers” in the United States will affect the demand for milk? Justify your answer.

**Exercise:****Problem:**

We know that a change in the price of a product causes a movement along the demand curve. Suppose consumers believe that prices will be rising in the future. How will that affect demand for the product in the present? Can you show this graphically?

**Exercise:****Problem:**

Suppose there is soda tax to curb obesity. What should a reduction in the soda tax do to the supply of sodas and to the equilibrium price and quantity? Can you show this graphically? *Hint:* assume that the soda tax is collected from the sellers

**Problems****Exercise:****Problem:**

[\[link\]](#) shows information on the demand and supply for bicycles, where the quantities of bicycles are measured in thousands.

Price	Qd	Qs
\$120	50	36
\$150	40	40
\$180	32	48
\$210	28	56
\$240	24	70

- a. What is the quantity demanded and the quantity supplied at a price of \$210?

- b. At what price is the quantity supplied equal to 48,000?
- c. Graph the demand and supply curve for bicycles. How can you determine the equilibrium price and quantity from the graph? How can you determine the equilibrium price and quantity from the table? What are the equilibrium price and equilibrium quantity?
- d. If the price was \$120, what would the quantities demanded and supplied be? Would a shortage or surplus exist? If so, how large would the shortage or surplus be?

### **Exercise:**

#### **Problem:**

The computer market in recent years has seen many more computers sell at much lower prices. What shift in demand or supply is most likely to explain this outcome? Sketch a demand and supply diagram and explain your reasoning for each.

- a. A rise in demand
- b. A fall in demand
- c. A rise in supply
- d. A fall in supply

### **References**

Landsburg, Steven E. *The Armchair Economist: Economics and Everyday Life*. New York: The Free Press. 2012. specifically Section IV: How Markets Work.

National Chicken Council. 2015. "Per Capita Consumption of Poultry and Livestock, 1965 to Estimated 2015, in Pounds." Accessed April 13, 2015. <http://www.nationalchickencouncil.org/about-the-industry/statistics/per-capita-consumption-of-poultry-and-livestock-1965-to-estimated-2012-in-pounds/>.

Wessel, David. "Saudi Arabia Fears \$40-a-Barrel Oil, Too." *The Wall Street Journal*. May 27, 2004, p. 42.  
<http://online.wsj.com/news/articles/SB108561000087822300>.

## **Glossary**

ceteris paribus

other things being equal

complements

goods that are often used together so that consumption of one good tends to enhance consumption of the other

factors of production

the combination of labor, materials, and machinery that is used to produce goods and services; also called inputs

inferior good

a good in which the quantity demanded falls as income rises, and in which quantity demanded rises and income falls

inputs

the combination of labor, materials, and machinery that is used to produce goods and services; also called factors of production

normal good

a good in which the quantity demanded rises as income rises, and in which quantity demanded falls as income falls

shift in demand

when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply

when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

substitute

a good that can replace another to some extent, so that greater consumption of one good can mean less of the other

## Changes in Equilibrium Price and Quantity: The Four-Step Process

By the end of this section, you will be able to:

- Identify equilibrium price and quantity through the four-step process
- Graph equilibrium price and quantity
- Contrast shifts of demand or supply and movements along a demand or supply curve
- Graph demand and supply curves, including equilibrium price and quantity, based on real-world examples

Let's begin this discussion with a single economic event. It might be an event that affects demand, like a change in income, population, tastes, prices of substitutes or complements, or expectations about future prices. It might be an event that affects supply, like a change in natural conditions, input prices, or technology, or government policies that affect production. How does this economic event affect equilibrium price and quantity? We will analyze this question using a four-step process.

Step 1. Draw a demand and supply model before the economic change took place. To establish the model requires four standard pieces of information: The law of demand, which tells us the slope of the demand curve; the law of supply, which gives us the slope of the supply curve; the shift variables for demand; and the shift variables for supply. From this model, find the initial equilibrium values for price and quantity.

Step 2. Decide whether the economic change being analyzed affects demand or supply. In other words, does the event refer to something in the list of demand factors or supply factors?

Step 3. Decide whether the effect on demand or supply causes the curve to shift to the right or to the left, and sketch the new demand or supply curve on the diagram. In other words, does the event increase or decrease the amount consumers want to buy or producers want to sell?

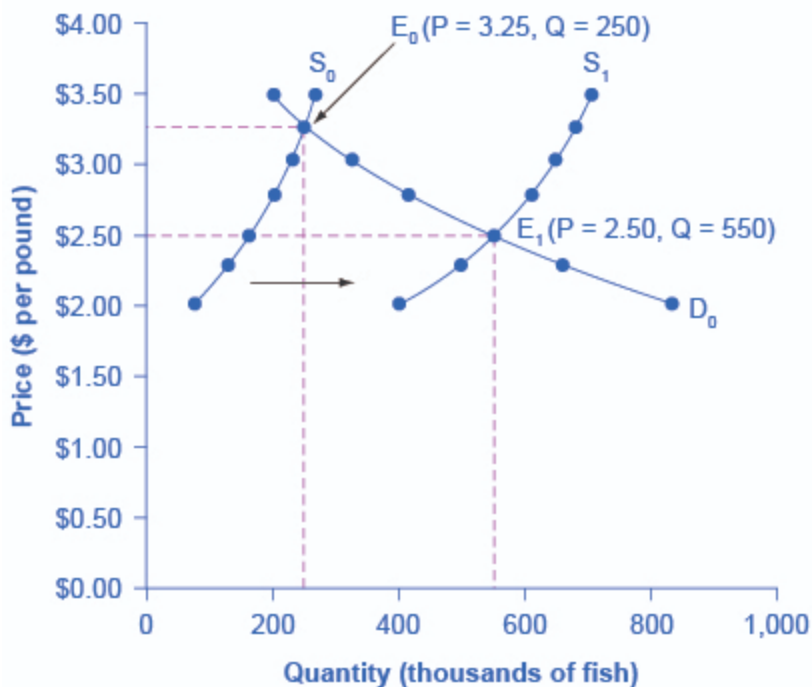
Step 4. Identify the new equilibrium and then compare the original equilibrium price and quantity to the new equilibrium price and quantity.

Let's consider one example that involves a shift in supply and one that involves a shift in demand. Then we will consider an example where both supply and demand shift.

## Good Weather for Salmon Fishing

In the summer of 2000, weather conditions were excellent for commercial salmon fishing off the California coast. Heavy rains meant higher than normal levels of water in the rivers, which helps the salmon to breed. Slightly cooler ocean temperatures stimulated the growth of plankton, the microscopic organisms at the bottom of the ocean food chain, providing everything in the ocean with a hearty food supply. The ocean stayed calm during fishing season, so commercial fishing operations did not lose many days to bad weather. How did these climate conditions affect the quantity and price of salmon? [\[link\]](#) illustrates the four-step approach, which is explained below, to work through this problem. [\[link\]](#) provides the information to work the problem as well.

### Good Weather for Salmon Fishing: The Four-Step Process



Unusually good weather leads to changes in the price and quantity of salmon.

<b>Price per Pound</b>	<b>Quantity Supplied in 1999</b>	<b>Quantity Supplied in 2000</b>	<b>Quantity Demanded</b>
\$2.00	80	400	840
\$2.25	120	480	680
\$2.50	160	550	550
\$2.75	200	600	450
\$3.00	230	640	350
\$3.25	250	670	250
\$3.50	270	700	200

### Salmon Fishing

Step 1. Draw a demand and supply model to illustrate the market for salmon in the year before the good weather conditions began. The demand curve  $D_0$  and the supply curve  $S_0$  show that the original equilibrium price is \$3.25 per pound and the original equilibrium quantity is 250,000 fish. (This price per pound is what commercial buyers pay at the fishing docks; what consumers pay at the grocery is higher.)

Step 2. Did the economic event affect supply or demand? Good weather is an example of a natural condition that affects supply.

Step 3. Was the effect on supply an increase or a decrease? Good weather is a change in natural conditions that increases the quantity supplied at any given price. The supply curve shifts to the right, moving from the original supply curve  $S_0$  to the new supply curve  $S_1$ , which is shown in both the table and the figure.

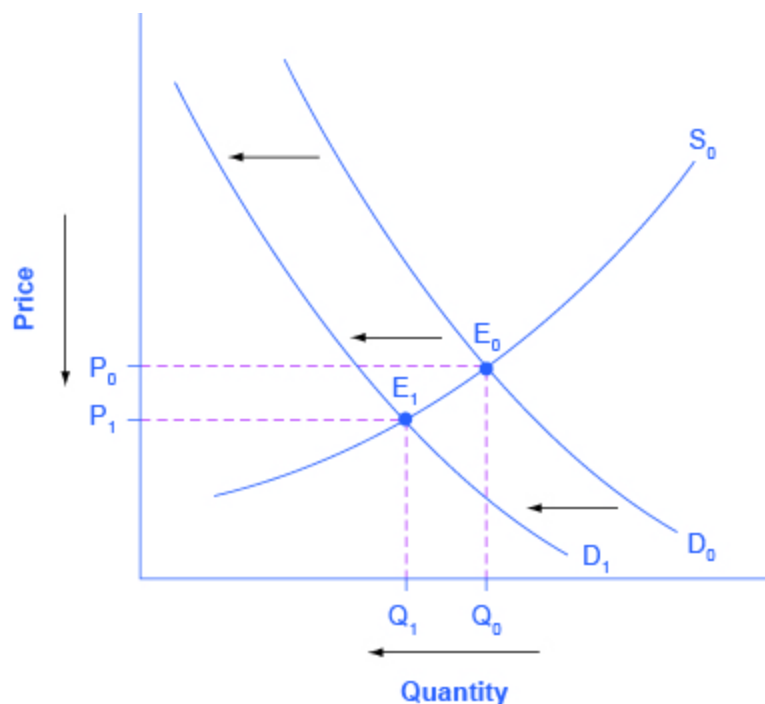
Step 4. Compare the new equilibrium price and quantity to the original equilibrium. At the new equilibrium  $E_1$ , the equilibrium price falls from \$3.25 to \$2.50, but the equilibrium quantity increases from 250,000 to 550,000 salmon. Notice that the equilibrium quantity demanded increased, even though the demand curve did not move.

In short, good weather conditions increased supply of the California commercial salmon. The result was a higher equilibrium quantity of salmon bought and sold in the market at a lower price.

## **Newspapers and the Internet**

According to the Pew Research Center for People and the Press, more and more people, especially younger people, are getting their news from online and digital sources. The majority of U.S. adults now own smartphones or tablets, and most of those Americans say they use them in part to get the news. From 2004 to 2012, the share of Americans who reported getting their news from digital sources increased from 24% to 39%. How has this affected consumption of print news media, and radio and television news? [\[link\]](#) and the text below illustrates using the four-step analysis to answer this question.

**The Print News Market: A Four-Step Analysis**



A change in tastes from print news sources to digital sources results in a leftward shift in demand for the former. The result is a decrease in both equilibrium price and quantity.

Step 1. Develop a demand and supply model to think about what the market looked like before the event. The demand curve  $D_0$  and the supply curve  $S_0$  show the original relationships. In this case, the analysis is performed without specific numbers on the price and quantity axis.

Step 2. Did the change described affect supply or demand? A change in tastes, from traditional news sources (print, radio, and television) to digital sources, caused a change in demand for the former.

Step 3. Was the effect on demand positive or negative? A shift to digital news sources will tend to mean a lower quantity demanded of traditional news sources at every given price, causing the demand curve for print and other traditional news sources to shift to the left, from  $D_0$  to  $D_1$ .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium ( $E_1$ ) occurs at a lower quantity and a lower price than the original equilibrium ( $E_0$ ).

The decline in print news reading predates 2004. Print newspaper circulation peaked in 1973 and has declined since then due to competition from television and radio news. In 1991, 55% of Americans indicated they got their news from print sources, while only 29% did so in 2012. Radio news has followed a similar path in recent decades, with the share of Americans getting their news from radio declining from 54% in 1991 to 33% in 2012. Television news has held its own over the last 15 years, with a market share staying in the mid to upper fifties. What does this suggest for the future, given that two-thirds of Americans under 30 years old say they do not get their news from television at all?

## **The Interconnections and Speed of Adjustment in Real Markets**

In the real world, many factors that affect demand and supply can change all at once. For example, the demand for cars might increase because of rising incomes and population, and it might decrease because of rising gasoline prices (a complementary good). Likewise, the supply of cars might increase because of innovative new technologies that reduce the cost of car production, and it might decrease as a result of new government regulations requiring the installation of costly pollution-control technology.

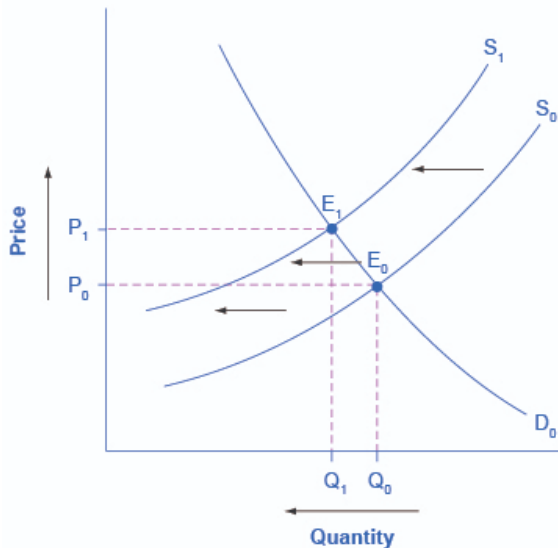
Moreover, rising incomes and population or changes in gasoline prices will affect many markets, not just cars. How can an economist sort out all these interconnected events? The answer lies in the *ceteris paribus* assumption. Look at how each economic event affects each market, one event at a time, holding all else constant. Then combine the analyses to see the net effect.

## **A Combined Example**

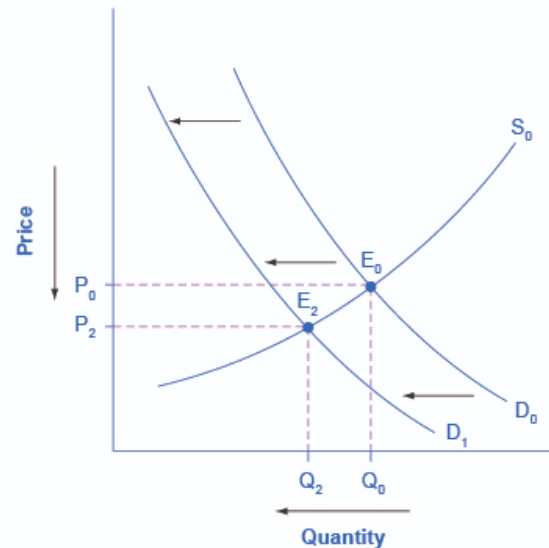
The U.S. Postal Service is facing difficult challenges. Compensation for postal workers tends to increase most years due to cost-of-living increases.

At the same time, more and more people are using email, text, and other digital message forms such as Facebook and Twitter to communicate with friends and others. What does this suggest about the continued viability of the Postal Service? [\[link\]](#) and the text below illustrates using the four-step analysis to answer this question.

### Higher Compensation for Postal Workers: A Four-Step Analysis



(a) Shift in supply



(b) Shift in demand

(a) Higher labor compensation causes a leftward shift in the supply curve, a decrease in the equilibrium quantity, and an increase in the equilibrium price. (b) A change in tastes away from Postal Services causes a leftward shift in the demand curve, a decrease in the equilibrium quantity, and a decrease in the equilibrium price.

Since this problem involves two disturbances, we need two four-step analyses, the first to analyze the effects of higher compensation for postal workers, the second to analyze the effects of many people switching from “snailmail” to email and other digital messages.

[\[link\]](#) (a) shows the shift in supply discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for the U.S. Postal Service looked like before this scenario starts. The demand

curve  $D_0$  and the supply curve  $S_0$  show the original relationships.

Step 2. Did the change described affect supply or demand? Labor compensation is a cost of production. A change in production costs caused a change in supply for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every given price, causing the supply curve for postal services to shift to the left, from  $S_0$  to  $S_1$ .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium ( $E_1$ ) occurs at a lower quantity and a higher price than the original equilibrium ( $E_0$ ).

[\[link\]](#) (b) shows the shift in demand discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for U.S. Postal Services looked like before this scenario starts. The demand curve  $D_0$  and the supply curve  $S_0$  show the original relationships. Note that this diagram is independent from the diagram in panel (a).

Step 2. Did the change described affect supply or demand? A change in tastes away from snailmail toward digital messages will cause a change in demand for the Postal Service.

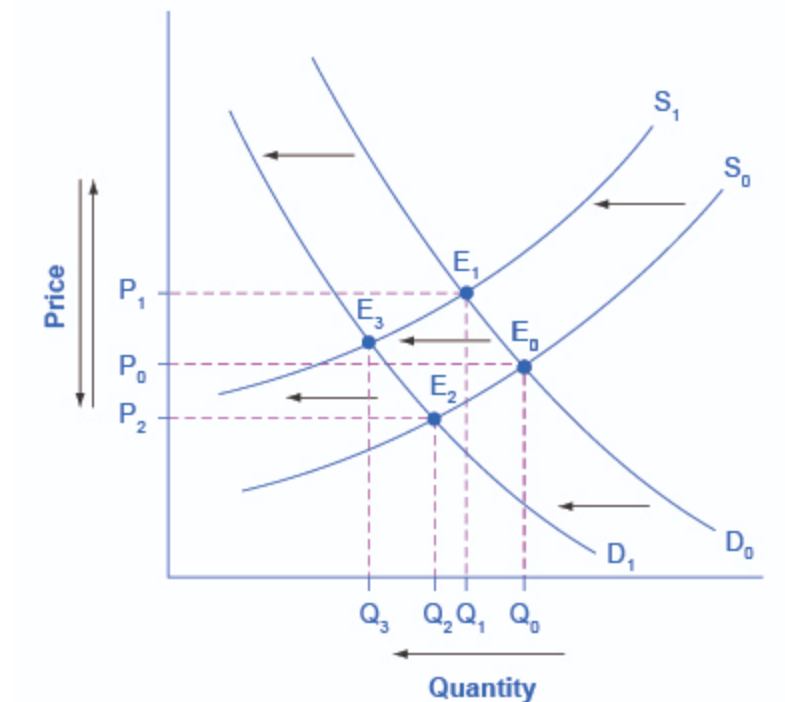
Step 3. Was the effect on demand positive or negative? A change in tastes away from snailmail toward digital messages causes lower quantity demanded of postal services at every given price, causing the demand curve for postal services to shift to the left, from  $D_0$  to  $D_1$ .

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium ( $E_2$ ) occurs at a lower quantity and a lower price than the original equilibrium ( $E_0$ ).

The final step in a scenario where both supply and demand shift is to combine the two individual analyses to determine what happens to the

equilibrium quantity and price. Graphically, we superimpose the previous two diagrams one on top of the other, as in [\[link\]](#).

### Combined Effect of Decreased Demand and Decreased Supply



Supply and demand shifts cause changes in equilibrium price and quantity.

Following are the results:

**Effect on Quantity:** The effect of higher labor compensation on Postal Services because it raises the cost of production is to decrease the equilibrium quantity. The effect of a change in tastes away from snailmail is to decrease the equilibrium quantity. Since both shifts are to the left, the overall impact is a decrease in the equilibrium quantity of Postal Services ( $Q_3$ ). This is easy to see graphically, since  $Q_3$  is to the left of  $Q_0$ .

**Effect on Price:** The overall effect on price is more complicated. The effect of higher labor compensation on Postal Services, because it raises the cost of production, is to increase the equilibrium price. The effect of a change in

tastes away from snailmail is to decrease the equilibrium price. Since the two effects are in opposite directions, unless we know the magnitudes of the two effects, the overall effect is unclear. This is not unusual. When both curves shift, typically we can determine the overall effect on price or on quantity, but not on both. In this case, we determined the overall effect on the equilibrium quantity, but not on the equilibrium price. In other cases, it might be the opposite.

The next Clear It Up feature focuses on the difference between shifts of supply or demand and movements along a curve.

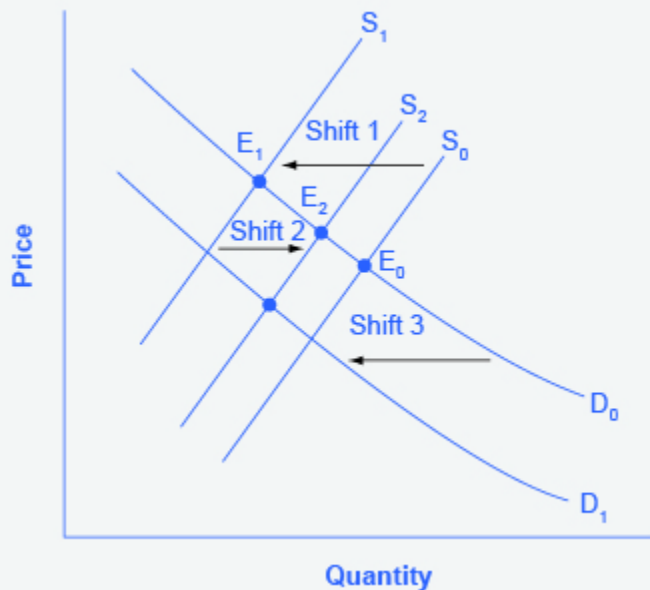
**Note:**

What is the difference between shifts of demand or supply versus movements along a demand or supply curve?

One common mistake in applying the demand and supply framework is to confuse the shift of a demand or a supply curve with movement along a demand or supply curve. As an example, consider a problem that asks whether a drought will increase or decrease the equilibrium quantity and equilibrium price of wheat. Lee, a student in an introductory economics class, might reason:

“Well, it is clear that a drought reduces supply, so I will shift back the supply curve, as in the shift from the original supply curve  $S_0$  to  $S_1$  shown on the diagram (called Shift 1). So the equilibrium moves from  $E_0$  to  $E_1$ , the equilibrium quantity is lower and the equilibrium price is higher. Then, a higher price makes farmers more likely to supply the good, so the supply curve shifts right, as shown by the shift from  $S_1$  to  $S_2$ , on the diagram (shown as Shift 2), so that the equilibrium now moves from  $E_1$  to  $E_2$ . The higher price, however, also reduces demand and so causes demand to shift back, like the shift from the original demand curve,  $D_0$  to  $D_1$  on the diagram (labeled Shift 3), and the equilibrium moves from  $E_2$  to  $E_3$ .”

Shifts of Demand or Supply versus Movements along a Demand or Supply Curve



A shift in one curve never causes a shift in the other curve. Rather, a shift in one curve causes a movement along the second curve.

At about this point, Lee suspects that this answer is headed down the wrong path. Think about what might be wrong with Lee's logic, and then read the answer that follows.

*Answer:* Lee's first step is correct: that is, a drought shifts back the supply curve of wheat and leads to a prediction of a lower equilibrium quantity and a higher equilibrium price. This corresponds to a movement along the original demand curve ( $D_0$ ), from  $E_0$  to  $E_1$ . The rest of Lee's argument is wrong, because it mixes up shifts in supply with quantity supplied, and shifts in demand with quantity demanded. A higher or lower price never shifts the supply curve, as suggested by the shift in supply from  $S_1$  to  $S_2$ . Instead, a price change leads to a movement along a given supply curve. Similarly, a higher or lower price never shifts a demand curve, as suggested in the shift from  $D_0$  to  $D_1$ . Instead, a price change leads to a movement along a given demand curve. Remember, a change in the price of a good never causes the demand or supply curve for that good to shift. Think carefully about the timeline of events: What happens first, what happens next? What is cause, what is effect? If you keep the order right, you are more likely to get the analysis correct.

In the four-step analysis of how economic events affect equilibrium price and quantity, the movement from the old to the new equilibrium seems immediate. As a practical matter, however, prices and quantities often do not zoom straight to equilibrium. More realistically, when an economic event causes demand or supply to shift, prices and quantities set off in the general direction of equilibrium. Indeed, even as they are moving toward one new equilibrium, prices are often then pushed by another change in demand or supply toward another equilibrium.

## Key Concepts and Summary

When using the supply and demand framework to think about how an event will affect the equilibrium price and quantity, proceed through four steps: (1) sketch a supply and demand diagram to think about what the market looked like before the event; (2) decide whether the event will affect supply or demand; (3) decide whether the effect on supply or demand is negative or positive, and draw the appropriate shifted supply or demand curve; (4) compare the new equilibrium price and quantity to the original ones.

## Self-Check Questions

### Exercise:

#### Problem:

Let's think about the market for air travel. From August 2014 to January 2015, the price of jet fuel decreased roughly 47%. Using the four-step analysis, how do you think this fuel price decrease affected the equilibrium price and quantity of air travel?

---

#### Solution:

Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? Jet fuel is a cost of producing air travel, so an increase in jet fuel price affects

supply.

Step 3. An increase in the price of jet fuel caused a decrease in the cost of air travel. We show this as a downward or rightward shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price of air travel and increasing the equilibrium quantity.

### **Exercise:**

#### **Problem:**

A tariff is a tax on imported goods. Suppose the U.S. government cuts the tariff on imported flat screen televisions. Using the four-step analysis, how do you think the tariff reduction will affect the equilibrium price and quantity of flat screen TVs?

---

#### **Solution:**

Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? A tariff is treated like a cost of production, so this affects supply.

Step 3. A tariff reduction is equivalent to a decrease in the cost of production, which we can show as a rightward (or downward) shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price and raising the equilibrium quantity.

### **Review Questions**

#### **Exercise:**

**Problem:**

How does one analyze a market where both demand and supply shift?

**Exercise:****Problem:**

What causes a movement along the demand curve? What causes a movement along the supply curve?

**Critical Thinking Questions****Exercise:****Problem:**

Use the four-step process to analyze the impact of the advent of the iPod (or other portable digital music players) on the equilibrium price and quantity of the Sony Walkman (or other portable audio cassette players).

**Exercise:****Problem:**

Use the four-step process to analyze the impact of a reduction in tariffs on imports of iPods on the equilibrium price and quantity of Sony Walkman-type products.

**Exercise:****Problem:**

Suppose both of these events took place at the same time. Combine your analyses of the impacts of the iPod and the tariff reduction to determine the likely impact on the equilibrium price and quantity of Sony Walkman-type products. Show your answer graphically.

**Problems**

**Exercise:****Problem:**

Demand and supply in the market for cheddar cheese is illustrated in [\[link\]](#). Graph the data and find the equilibrium. Next, create a table showing the change in quantity demanded or quantity supplied, and a graph of the new equilibrium, in each of the following situations:

- a. The price of milk, a key input for cheese production, rises, so that the supply decreases by 80 pounds at every price.
- b. A new study says that eating cheese is good for your health, so that demand increases by 20% at every price.

Price per Pound	Qd	Qs
\$3.00	750	540
\$3.20	700	600
<b>\$3.40</b>	<b>650</b>	<b>650</b>
\$3.60	620	700
\$3.80	600	720
\$4.00	590	730

**Exercise:**

**Problem:**

Supply and demand for movie tickets in a city are shown in [\[link\]](#). Graph demand and supply and identify the equilibrium. Then calculate in a table and graph the effect of the following two changes.

- a. Three new nightclubs open. They offer decent bands and have no cover charge, but make their money by selling food and drink. As a result, demand for movie tickets falls by six units at every price.
- b. The city eliminates a tax that it had been placing on all local entertainment businesses. The result is that the quantity supplied of movies at any given price increases by 10%.

Price per Pound	Qd	Qs
\$5.00	26	16
\$6.00	24	18
\$7.00	22	20
\$8.00	21	21
\$9.00	20	22

**References**

Pew Research Center. "Pew Research: Center for the People & the Press."  
<http://www.people-press.org/>.

## Price Ceilings and Price Floors

By the end of this section, you will be able to:

- Explain price controls, price ceilings, and price floors
- Analyze demand and supply as a social adjustment mechanism

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing “too high” or falling “too low.”

The demand and supply model shows how people and firms will react to the incentives provided by these laws to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

### Price Ceilings

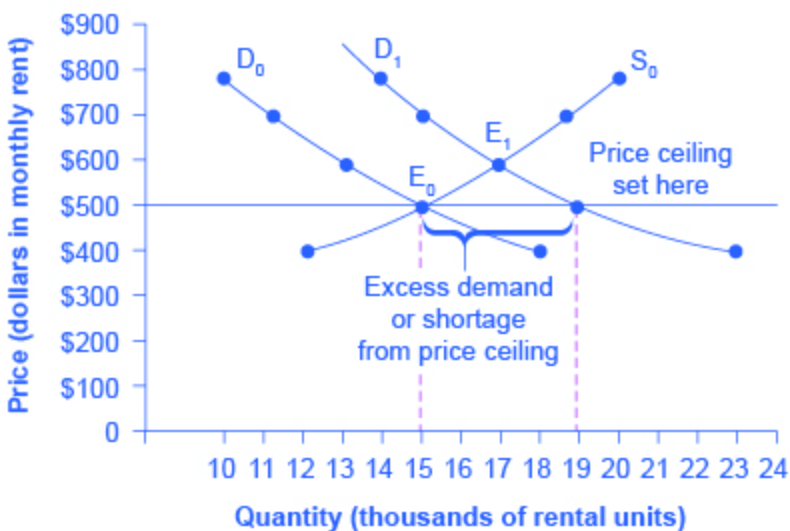
Laws that government enacts to regulate prices are called **Price controls**. Price controls come in two flavors. A **price ceiling** keeps a price from rising above a certain level (the “ceiling”), while a **price floor** keeps a price from falling below a certain level (the “floor”). This section uses the demand and supply framework to analyze price ceilings. The next section discusses price floors.

In many markets for goods and services, demanders outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. In some cities, such as Albany, renters have pressed political leaders to pass rent control laws, a price ceiling that usually works by stating that rents can be raised by only a certain maximum percentage each year.

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps

locally-based businesses expand, bringing higher incomes and more people into the area. Changes of this sort can cause a change in the demand for rental housing, as [\[link\]](#) illustrates. The original equilibrium ( $E_0$ ) lies at the intersection of supply curve  $S_0$  and demand curve  $D_0$ , corresponding to an equilibrium price of \$500 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental housing to the right, as shown by the data in [\[link\]](#) and the shift from  $D_0$  to  $D_1$  on the graph. In this market, at the new equilibrium  $E_1$ , the price of a rental unit would rise to \$600 and the equilibrium quantity would increase to 17,000 units.

### A Price Ceiling Example—Rent Control



The original intersection of demand and supply occurs at  $E_0$ . If demand shifts from  $D_0$  to  $D_1$ , the new equilibrium would be at  $E_1$ —unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the change in demand, the quantity demanded rises to 19,000, resulting in a shortage.

<b>Price</b>	<b>Original Quantity Supplied</b>	<b>Original Quantity Demanded</b>	<b>New Quantity Demanded</b>
\$400	12,000	18,000	23,000
\$500	15,000	15,000	19,000
\$600	17,000	13,000	17,000
\$700	19,000	11,000	15,000
\$800	20,000	10,000	14,000

## Rent Control

Suppose that a rent control law is passed to keep the price at the original equilibrium of \$500 for a typical apartment. In [\[link\]](#), the horizontal line at the price of \$500 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price (\$500), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a shortage of rental housing. One of the ironies of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of \$600 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or potential renters) lose their housing as landlords convert apartments to co-ops and condos. Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. So if renters get “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings have been proposed for other products. For example, price ceilings to limit what producers can charge have been proposed in recent years for prescription drugs, doctor and hospital fees, the charges made by some automatic teller bank machines, and auto insurance rates. Price ceilings are enacted in an attempt to keep prices low for those who demand the product. But when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

## **Price Floors**

A price floor is the lowest legal price that can be paid in markets for goods and services, labor, or financial capital. Perhaps the best-known example of a price floor is the minimum wage, which is based on the normative view that someone working full time ought to be able to afford a basic standard of living. The federal minimum wage at the end of 2014 was \$7.25 per hour, which yields an income for a single person slightly higher than the poverty line. As the cost of living rises over time, the Congress periodically raises the federal minimum wage.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely. So even if, on average, farm incomes are adequate, some years they can be quite low. The purpose of price supports is to prevent these swings.

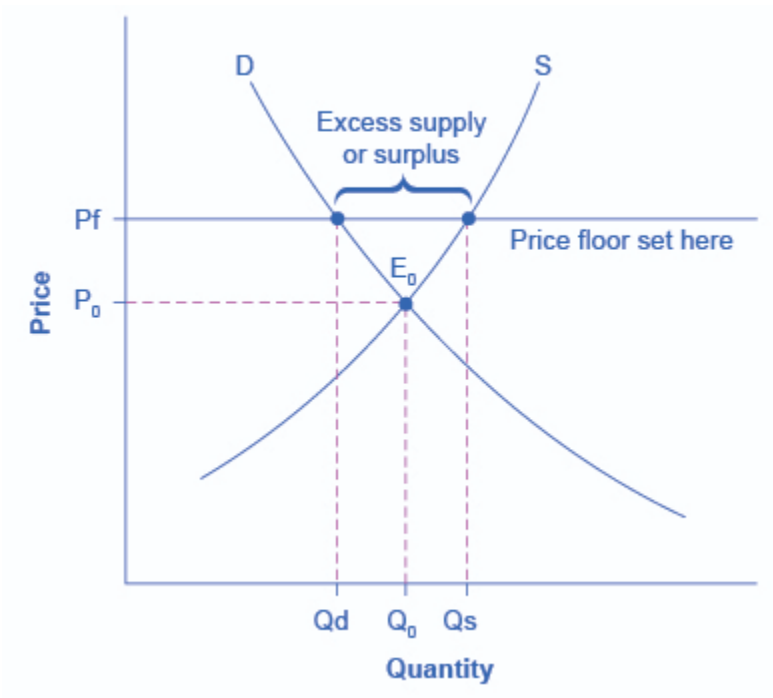
The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. According to the Common Agricultural Policy reform passed in 2013, the European Union (EU) will spend about 60 billion euros per year, or 67 billion dollars per year, or roughly 38% of the EU budget, on price supports for Europe’s farmers from 2014 to 2020.

[\[link\]](#) illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point  $E_0$ , with price  $P_0$  and quantity  $Q_0$ . However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price  $P_f$  shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity demanded ( $Q_d$ ). When quantity supplied exceeds quantity demanded, a surplus exists.

The high-income areas of the world, including the United States, Europe, and Japan, are estimated to spend roughly \$1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Numerous proposals have been offered for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. Either because this is viewed by the population as supporting the traditional rural way of life or because of the lobbying power of the agribusiness industry.

For more detail on the effects price ceilings and floors have on demand and supply, see the following Clear It Up feature.

**European Wheat Prices: A Price Floor Example**



The intersection of demand (D) and supply (S) would be at the equilibrium point  $E_0$ .

However, a price floor set at  $P_f$  holds the price above  $E_0$  and prevents it from falling.

The result of the price floor is that the quantity supplied  $Q_s$  exceeds the quantity demanded  $Q_d$ . There is excess supply, also called a surplus.

**Note:**

Do price ceilings and floors change demand or supply?

Neither price ceilings nor price floors cause demand or supply to change. They simply set a price that limits what can be legally charged in the market. Remember, changes in price do not cause demand or supply to change. Price ceilings and price floors can cause a different choice of quantity demanded along a demand curve, but they do not move the

demand curve. Price controls can cause a different choice of quantity supplied along a supply curve, but they do not shift the supply curve.

## Key Concepts and Summary

Price ceilings prevent a price from rising above a certain level. When a price ceiling is set below the equilibrium price, quantity demanded will exceed quantity supplied, and excess demand or shortages will result. Price floors prevent a price from falling below a certain level. When a price floor is set above the equilibrium price, quantity supplied will exceed quantity demanded, and excess supply or surpluses will result. Price floors and price ceilings often lead to unintended consequences.

## Self-Check Questions

### Exercise:

#### Problem:

What is the effect of a price ceiling on the quantity demanded of the product? What is the effect of a price ceiling on the quantity supplied? Why exactly does a price ceiling cause a shortage?

---

#### Solution:

A price ceiling (which is below the equilibrium price) will cause the quantity demanded to rise and the quantity supplied to fall. This is why a price ceiling creates a shortage.

### Exercise:

**Problem:** Does a price ceiling change the equilibrium price?

---

#### Solution:

A price ceiling is just a legal restriction. Equilibrium is an economic condition. People may or may not obey the price ceiling, so the actual price may be at or above the price ceiling, but the price ceiling does not change the equilibrium price.

**Exercise:**

**Problem:**

What would be the impact of imposing a price floor below the equilibrium price?

---

**Solution:**

A price ceiling is a legal maximum price, but a price floor is a legal minimum price and, consequently, it would leave room for the price to rise to its equilibrium level. In other words, a price floor below equilibrium will not be binding and will have no effect.

## Review Questions

**Exercise:**

**Problem:**

Does a price ceiling attempt to make a price higher or lower?

**Exercise:**

**Problem:**

How does a price ceiling set below the equilibrium level affect quantity demanded and quantity supplied?

**Exercise:**

**Problem:** Does a price floor attempt to make a price higher or lower?

**Exercise:**

**Problem:**

How does a price floor set above the equilibrium level affect quantity demanded and quantity supplied?

**Critical Thinking Questions****Exercise:****Problem:**

Most government policy decisions have winners and losers. What are the effects of raising the minimum wage? It is more complex than simply producers lose and workers gain. Who are the winners and who are the losers, and what exactly do they win and lose? To what extent does the policy change achieve its goals?

**Exercise:****Problem:**

Agricultural price supports result in governments holding large inventories of agricultural products. Why do you think the government cannot simply give the products away to poor people?

**Exercise:****Problem:**

Can you propose a policy that would induce the market to supply more rental housing units?

**Problems****Exercise:**

### Problem:

A low-income country decides to set a price ceiling on bread so it can make sure that bread is affordable to the poor. The conditions of demand and supply are given in [\[link\]](#). What are the equilibrium price and equilibrium quantity before the price ceiling? What will the excess demand or the shortage (that is, quantity demanded minus quantity supplied) be if the price ceiling is set at \$2.40? At \$2.00? At \$3.60?

Price	Qd	Qs
\$1.60	9,000	5,000
\$2.00	8,500	5,500
\$2.40	8,000	6,400
<b>\$2.80</b>	<b>7,500</b>	<b>7,500</b>
\$3.20	7,000	9,000
\$3.60	6,500	11,000
\$4.00	6,000	15,000

### Glossary

price ceiling

a legal maximum price

price control

government laws to regulate prices instead of letting market forces determine prices

price floor

a legal minimum price

total surplus

see social surplus

## Demand, Supply, and Efficiency

By the end of this section, you will be able to:

- Contrast consumer surplus, producer surplus, and social surplus
- Explain why price floors and price ceilings can be inefficient
- Analyze demand and supply as a social adjustment mechanism

The familiar demand and supply diagram holds within it the concept of economic efficiency. One typical way that economists define efficiency is when it is impossible to improve the situation of one party without imposing a cost on another. Conversely, if a situation is inefficient, it becomes possible to benefit at least one party without imposing costs on others.

Efficiency in the demand and supply model has the same basic meaning: The economy is getting as much benefit as possible from its scarce resources and all the possible gains from trade have been achieved. In other words, the optimal amount of each good and service is being produced and consumed.

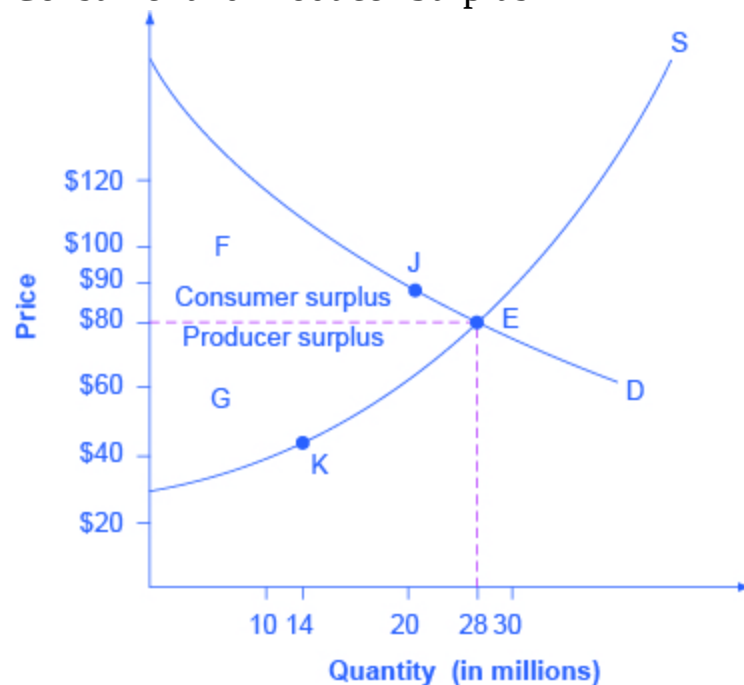
## Consumer Surplus, Producer Surplus, Social Surplus

Consider a market for tablet computers, as shown in [\[link\]](#). The equilibrium price is \$80 and the equilibrium quantity is 28 million. To see the benefits to consumers, look at the segment of the demand curve above the equilibrium point and to the left. This portion of the demand curve shows that at least some demanders would have been willing to pay more than \$80 for a tablet.

For example, point J shows that if the price was \$90, 20 million tablets would be sold. Those consumers who would have been willing to pay \$90 for a tablet based on the utility they expect to receive from it, but who were able to pay the equilibrium price of \$80, clearly received a benefit beyond what they had to pay for. Remember, the demand curve traces consumers' willingness to pay for different quantities. The amount that individuals would have been willing to pay, minus the amount that they actually paid, is

called **consumer surplus**. Consumer surplus is the area labeled F—that is, the area above the market price and below the demand curve.

### Consumer and Producer Surplus



The somewhat triangular area labeled by F shows the area of consumer surplus, which shows that the equilibrium price in the market was less than what many of the consumers were willing to pay. Point J on the demand curve shows that, even at the price of \$90, consumers would have been willing to purchase a quantity of 20 million. The somewhat triangular area labeled by G shows the area of producer surplus, which shows that the equilibrium price received in the market was more than what many of the producers were willing to accept for their products. For example, point K on the supply curve shows that at a price of \$45, firms would have been willing to supply a quantity of 14 million.

The supply curve shows the quantity that firms are willing to supply at each price. For example, point K in [\[link\]](#) illustrates that, at \$45, firms would still have been willing to supply a quantity of 14 million. Those producers who would have been willing to supply the tablets at \$45, but who were instead able to charge the equilibrium price of \$80, clearly received an extra benefit beyond what they required to supply the product. The amount that a seller is paid for a good minus the seller's actual cost is called **producer surplus**. In [\[link\]](#), producer surplus is the area labeled G—that is, the area between the market price and the segment of the supply curve below the equilibrium.

The sum of consumer surplus and producer surplus is **social surplus**, also referred to as **economic surplus** or **total surplus**. In [\[link\]](#), social surplus would be shown as the area  $F + G$ . Social surplus is larger at equilibrium quantity and price than it would be at any other quantity. This demonstrates the economic efficiency of the market equilibrium. In addition, at the efficient level of output, it is impossible to produce greater consumer surplus without reducing producer surplus, and it is impossible to produce greater producer surplus without reducing consumer surplus.

## Inefficiency of Price Floors and Price Ceilings

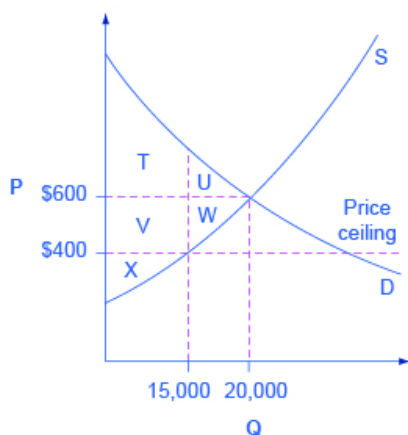
The imposition of a price floor or a price ceiling will prevent a market from adjusting to its equilibrium price and quantity, and thus will create an inefficient outcome. But there is an additional twist here. Along with creating inefficiency, price floors and ceilings will also transfer some consumer surplus to producers, or some producer surplus to consumers.

Imagine that several firms develop a promising but expensive new drug for treating back pain. If this therapy is left to the market, the equilibrium price will be \$600 per month and 20,000 people will use the drug, as shown in [\[link\]](#) (a). The original level of consumer surplus is  $T + U$  and producer surplus is  $V + W + X$ . However, the government decides to impose a price ceiling of \$400 to make the drug more affordable. At this price ceiling, firms in the market now produce only 15,000.

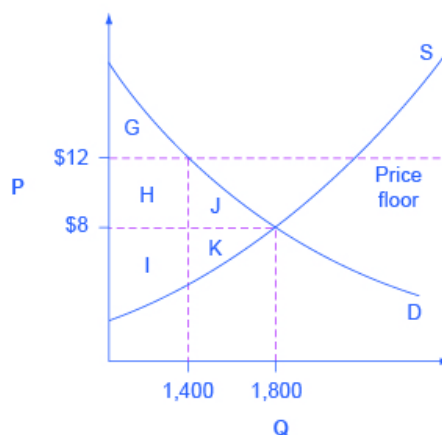
As a result, two changes occur. First, an inefficient outcome occurs and the total surplus of society is reduced. The loss in social surplus that occurs when the economy produces at an inefficient quantity is called **deadweight loss**. In a very real sense, it is like money thrown away that benefits no one. In [\[link\]](#) (a), the deadweight loss is the area U + W. When deadweight loss exists, it is possible for both consumer and producer surplus to be higher, in this case because the price control is blocking some suppliers and demanders from transactions they would both be willing to make.

A second change from the price ceiling is that some of the producer surplus is transferred to consumers. After the price ceiling is imposed, the new consumer surplus is T + V, while the new producer surplus is X. In other words, the price ceiling transfers the area of surplus (V) from producers to consumers. Note that the gain to consumers is less than the loss to producers, which is just another way of seeing the deadweight loss.

### Efficiency and Price Floors and Ceilings



(a) Reduced social surplus from a price ceiling



(b) Reduced social surplus from a price floor

- (a) The original equilibrium price is \$600 with a quantity of 20,000. Consumer surplus is T + U, and producer surplus is V + W + X. A price ceiling is imposed at \$400, so firms in the market now produce only a quantity of 15,000. As a result, the new consumer surplus is T + V, while the new producer surplus is X. (b) The original equilibrium is \$8 at a quantity of 1,800. Consumer surplus is G + H + J, and producer surplus is I + K. A price floor is imposed at \$12, which means that quantity demanded falls to 1,400. As a result, the new consumer surplus is G, and the new producer surplus is H + I.

[\[link\]](#) (b) shows a price floor example using a string of struggling movie theaters, all in the same city. The current equilibrium is \$8 per movie ticket, with 1,800 people attending movies. The original consumer surplus is  $G + H + J$ , and producer surplus is  $I + K$ . The city government is worried that movie theaters will go out of business, reducing the entertainment options available to citizens, so it decides to impose a price floor of \$12 per ticket. As a result, the quantity demanded of movie tickets falls to 1,400. The new consumer surplus is  $G$ , and the new producer surplus is  $H + I$ . In effect, the price floor causes the area  $H$  to be transferred from consumer to producer surplus, but also causes a deadweight loss of  $J + K$ .

This analysis shows that a price ceiling, like a law establishing rent controls, will transfer some producer surplus to consumers—which helps to explain why consumers often favor them. Conversely, a price floor like a guarantee that farmers will receive a certain price for their crops will transfer some consumer surplus to producers, which explains why producers often favor them. However, both price floors and price ceilings block some transactions that buyers and sellers would have been willing to make, and creates deadweight loss. Removing such barriers, so that prices and quantities can adjust to their equilibrium level, will increase the economy's social surplus.

## **Demand and Supply as a Social Adjustment Mechanism**

The demand and supply model emphasizes that prices are not set only by demand or only by supply, but by the interaction between the two. In 1890, the famous economist Alfred Marshall wrote that asking whether supply or demand determined a price was like arguing “whether it is the upper or the under blade of a pair of scissors that cuts a piece of paper.” The answer is that both blades of the demand and supply scissors are always involved.

The adjustments of equilibrium price and quantity in a market-oriented economy often occur without much government direction or oversight. If the coffee crop in Brazil suffers a terrible frost, then the supply curve of

coffee shifts to the left and the price of coffee rises. Some people—call them the coffee addicts—continue to drink coffee and pay the higher price. Others switch to tea or soft drinks. No government commission is needed to figure out how to adjust coffee prices, which companies will be allowed to process the remaining supply, which supermarkets in which cities will get how much coffee to sell, or which consumers will ultimately be allowed to drink the brew. Such adjustments in response to price changes happen all the time in a market economy, often so smoothly and rapidly that we barely notice them.

Think for a moment of all the seasonal foods that are available and inexpensive at certain times of the year, like fresh corn in midsummer, but more expensive at other times of the year. People alter their diets and restaurants alter their menus in response to these fluctuations in prices without fuss or fanfare. For both the U.S. economy and the world economy as a whole, markets—that is, demand and supply—are the primary social mechanism for answering the basic questions about what is produced, how it is produced, and for whom it is produced.

**Note:**

**Why Can We Not Get Enough of Organic?**

Organic food is grown without synthetic pesticides, chemical fertilizers or genetically modified seeds. In recent decades, the demand for organic products has increased dramatically. The Organic Trade Association reported sales increased from \$1 billion in 1990 to \$35.1 billion in 2013, more than 90% of which were sales of food products.

Why, then, are organic foods more expensive than their conventional counterparts? The answer is a clear application of the theories of supply and demand. As people have learned more about the harmful effects of chemical fertilizers, growth hormones, pesticides and the like from large-scale factory farming, our tastes and preferences for safer, organic foods have increased. This change in tastes has been reinforced by increases in income, which allow people to purchase pricier products, and has made organic foods more mainstream. This has led to an increased demand for organic foods. Graphically, the demand curve has shifted right, and we

have moved up the supply curve as producers have responded to the higher prices by supplying a greater quantity.

In addition to the movement along the supply curve, we have also had an increase in the number of farmers converting to organic farming over time. This is represented by a shift to the right of the supply curve. Since both demand and supply have shifted to the right, the resulting equilibrium quantity of organic foods is definitely higher, but the price will only fall when the increase in supply is larger than the increase in demand. We may need more time before we see lower prices in organic foods. Since the production costs of these foods may remain higher than conventional farming, because organic fertilizers and pest management techniques are more expensive, they may never fully catch up with the lower prices of non-organic foods.

As a final, specific example: The Environmental Working Group's "Dirty Dozen" list of fruits and vegetables, which test high for pesticide residue even after washing, was released in April 2013. The inclusion of strawberries on the list has led to an increase in demand for organic strawberries, resulting in both a higher equilibrium price and quantity of sales.

Consumer surplus is the gap between the price that consumers are willing to pay, based on their preferences, and the market equilibrium price. Producer surplus is the gap between the price for which producers are willing to sell a product, based on their costs, and the market equilibrium price. Social surplus is the sum of consumer surplus and producer surplus. Total surplus is larger at the equilibrium quantity and price than it will be at any other quantity and price. Deadweight loss is loss in total surplus that occurs when the economy produces at an inefficient quantity.

**Exercise:**

**Problem:**

Does a price ceiling increase or decrease the number of transactions in a market? Why? What about a price floor?

---

**Solution:**

Assuming that people obey the price ceiling, the market price will be below equilibrium, which means that  $Q_d$  will be more than  $Q_s$ . Buyers can only buy what is offered for sale, so the number of transactions will fall to  $Q_s$ . This is easy to see graphically. By analogous reasoning, with a price floor the market price will be above the equilibrium price, so  $Q_d$  will be less than  $Q_s$ . Since the limit on transactions here is demand, the number of transactions will fall to  $Q_d$ . Note that because both price floors and price ceilings reduce the number of transactions, social surplus is less.

**Exercise:**

**Problem:**

If a price floor benefits producers, why does a price floor reduce social surplus?

---

**Solution:**

Because the losses to consumers are greater than the benefits to producers, so the net effect is negative. Since the lost consumer surplus is greater than the additional producer surplus, social surplus falls.

**Exercise:**

**Problem:**

What is consumer surplus? How is it illustrated on a demand and supply diagram?

**Exercise:**

**Problem:**

What is producer surplus? How is it illustrated on a demand and supply diagram?

**Exercise:**

**Problem:**

What is total surplus? How is it illustrated on a demand and supply diagram?

**Exercise:**

**Problem:**

What is the relationship between total surplus and economic efficiency?

**Exercise:**

**Problem:** What is deadweight loss?

**Exercise:**

**Problem:**

What term would an economist use to describe what happens when a shopper gets a “good deal” on a product?

**Exercise:**

**Problem:** Explain why voluntary transactions improve social welfare.

**Exercise:**

**Problem:**

Why would a free market never operate at a quantity greater than the equilibrium quantity? *Hint:* What would be required for a transaction to occur at that quantity?

## Glossary

consumer surplus

the extra benefit consumers receive from buying a good or service, measured by what the individuals would have been willing to pay

minus the amount that they actually paid

deadweight loss

the loss in social surplus that occurs when a market produces an inefficient quantity

economic surplus

see social surplus

producer surplus

the extra benefit producers receive from selling a good or service, measured by the price the producer actually received minus the price the producer would have been willing to accept

social surplus

the sum of consumer surplus and producer surplus

## Introduction to Consumer Choices

class="introduction"

### Investment Choices

Higher education is generally viewed as a good investment, if one can afford it, regardless of the state of the economy.  
(Credit: modification of work by Jason Bache/Flickr  
r Creative Commons)

**Note:****"Eeny, Meeny, Miney, Moe"—Making Choices**

The Great Recession of 2008–2009 touched families around the globe. In too many countries, workers found themselves out of a job. In developed countries, unemployment compensation provided a safety net, but families still saw a marked decrease in disposable income and had to make tough spending decisions. Of course, non-essential, discretionary spending was the first to go.

Even so, there was one particular category that saw a universal increase in spending world-wide during that time—an 18% uptick in the United States, specifically. You might guess that consumers began eating more meals at home, increasing spending at the grocery store. But the Bureau of Labor Statistics' Consumer Expenditure Survey, which tracks U.S. food spending over time, showed “real total food spending by U.S. households declined five percent between 2006 and 2009.” So, it was not groceries. Just what product would people around the world demand more of during tough economic times, and more importantly, why? (Find out at chapter's end.)

That question leads us to this chapter's topic—analyzing how consumers make choices. For most consumers, using “eeny, meeny, miney, moe” is not how they make decisions; their decision-making processes have been educated far beyond a children's rhyme.

**Note:**

**Introduction to Consumer Choices**

In this chapter, you will learn about:

- Consumption Choices
- How Changes in Income and Prices Affect Consumption Choices
- Labor-Leisure Choices
- Intertemporal Choices in Financial Capital Markets

Microeconomics seeks to understand the behavior of individual economic agents such as individuals and businesses. Economists believe that individuals' decisions, such as what goods and services to buy, can be analyzed as choices made within certain budget constraints. Generally, consumers are trying to get the most for their limited budget. In economic terms they are trying to maximize total utility, or satisfaction, given their budget constraint.

Everyone has their own personal tastes and preferences. The French say: *Chacun à son goût*, or “Each to his own taste.” An old Latin saying states, *De gustibus non est disputandum* or “There's no disputing about taste.” If people's decisions are based on their own tastes and personal preferences, however, then how can economists hope to analyze the choices consumers make?

An economic explanation for why people make different choices begins with accepting the proverbial wisdom that tastes are a matter of personal preference. But economists also believe that the choices people make are influenced by their incomes, by the prices of goods and services they

consume, and by factors like where they live. This chapter introduces the economic theory of how consumers make choices about what to buy, how much to work, and how much to save.

The analysis in this chapter will build on the three budget constraints introduced in the [Choice in a World of Scarcity](#) chapter. These were the consumption choice budget constraint, the labor-leisure budget constraint, and the intertemporal budget constraint. This chapter will also illustrate how economic theory provides a tool to systematically look at the full range of possible consumption choices to predict how consumption responds to changes in prices or incomes. After reading this chapter, consult the appendix [Indifference Curves](#) to learn more about representing utility and choice through indifference curves.

## Consumption Choices

By the end of this section, you will be able to:

- Calculate total utility
- Propose decisions that maximize utility
- Explain marginal utility and the significance of diminishing marginal utility

Information on the consumption choices of Americans is available from the Consumer Expenditure Survey carried out by the U.S. Bureau of Labor Statistics. [\[link\]](#) shows spending patterns for the average U.S. household. The first row shows income and, after taxes and personal savings are subtracted, it shows that, in 2015, the average U.S. household spent \$48,109 on consumption. The table then breaks down consumption into various categories. The average U.S. household spent roughly one-third of its consumption on shelter and other housing expenses, another one-third on food and vehicle expenses, and the rest on a variety of items, as shown. Of course, these patterns will vary for specific households by differing levels of family income, by geography, and by preferences.

<b>Average Household Income before Taxes</b>	<b>\$62,481</b>
<b>Average Annual Expenditures</b>	<b>\$48.109</b>
Food at home	\$3,264
Food away from home	\$2,505
Housing	\$16,557
Apparel and services	\$1,700
Transportation	\$7,677
Healthcare	\$3,157
Entertainment	\$2,504
Education	\$1,074
Personal insurance and pensions	\$5,357
All else: alcohol, tobacco, reading, personal care, cash contributions, miscellaneous	\$3,356

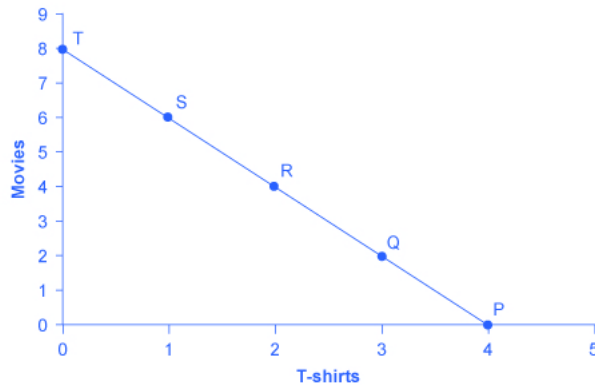
U.S. Consumption Choices in 2015(Source: <http://www.bls.gov/cex/csxann13.pdf>)

## Total Utility and Diminishing Marginal Utility

To understand how a household will make its choices, economists look at what consumers can afford, as shown in a **budget constraint line**, and the **total utility** or satisfaction derived from those choices. In a budget constraint line, the quantity of one good is measured on the horizontal axis and the quantity of the other good is measured on the vertical axis. The budget constraint line shows the various combinations of two goods that are affordable given consumer income. Consider the situation of José, shown in [\[link\]](#). José likes to collect T-shirts and watch movies.

In [\[link\]](#), the quantity of T-shirts is shown on the horizontal axis, while the quantity of movies is shown on the vertical axis. If José had unlimited income or goods were free, then he could consume without limit. But José, like all of us, faces a budget constraint. José has a total of \$56 to spend. The price of T-shirts is \$14 and the price of movies is \$7. Notice that the vertical intercept of the budget constraint line is at eight movies and zero T-shirts ( $\$56/\$7=8$ ). The horizontal intercept of the budget constraint is four, where José spends all of his money on T-shirts and no movies ( $\$56/\$14=4$ ). The slope of the budget constraint line is rise/run or  $-8/4=-2$ . The specific choices along the budget constraint line show the combinations of T-shirts and movies that are affordable.

#### A Choice between Consumption Goods



José has income of \$56. Movies cost \$7 and T-shirts cost \$14. The points on the budget constraint line show the combinations of movies and T-shirts that are affordable.

José wishes to choose the combination that will provide him with the greatest utility, which is the term economists use to describe a person's level of satisfaction or happiness with his or her choices.

Let's begin with an assumption, which will be discussed in more detail later, that José can measure his own utility with something called *utils*. (It is important to note that you cannot make comparisons between the utils of individuals; if one person gets 20 utils from a cup of coffee and another gets 10 utils, this does not mean that the first person gets more enjoyment from the coffee than the other or that they enjoy the coffee twice as much.) [\[link\]](#) shows how José's utility is connected with his consumption of T-shirts or movies. The first column of the table shows the quantity of T-shirts consumed. The second column shows the total utility, or total amount of satisfaction, that José receives from consuming that number of T-shirts. The most common pattern of total utility, as shown here, is that consuming additional goods leads to greater total utility, but at a decreasing rate. The third column shows **marginal utility**, which is the additional utility provided by one additional unit of consumption. This equation for marginal utility is:

**Equation:**

$$MU = \frac{\text{change in total utility}}{\text{change in quantity}}$$

Notice that marginal utility diminishes as additional units are consumed, which means that each subsequent unit of a good consumed provides less *additional* utility. For example, the first T-shirt José picks is his favorite and it gives him an addition of 22 utils. The fourth T-shirt is just to something to wear when all his other clothes are in the wash and yields only 18 additional utils. This is an example of the law of **diminishing marginal utility**, which holds that the additional utility decreases with each unit added.

The rest of [\[link\]](#) shows the quantity of movies that José attends, and his total and marginal utility from seeing each movie. Total utility follows the expected pattern: it increases as the number of movies seen rises. Marginal utility also follows the expected pattern: each additional movie brings a smaller gain in utility than the previous one. The first movie José attends is the one he wanted to see the most, and thus provides him with the highest level

of utility or satisfaction. The fifth movie he attends is just to kill time. Notice that total utility is also the sum of the marginal utilities. Read the next Work It Out feature for instructions on how to calculate total utility.

T-Shirts (Quantity)	Total Utility	Marginal Utility	Movies (Quantity)	Total Utility	Marginal Utility
1	22	22	1	16	16
2	43	21	2	31	15
3	63	20	3	45	14
4	81	18	4	58	13
5	97	16	5	70	12
6	111	14	6	81	11
7	123	12	7	91	10
8	133	10	8	100	9

#### Total and Marginal Utility

[\[link\]](#) looks at each point on the budget constraint in [\[link\]](#), and adds up José's total utility for five possible combinations of T-shirts and movies.

Point	T-Shirts	Movies	Total Utility
P	4	0	$81 + 0 = 81$
Q	3	2	$63 + 31 = 94$
R	2	4	$43 + 58 = 101$
S	1	6	$22 + 81 = 103$
T	0	8	$0 + 100 = 100$

#### Finding the Choice with the Highest Utility

##### Note:

##### Calculating Total Utility

Let's look at how José makes his decision in more detail.

Step 1. Observe that, at point Q (for example), José consumes three T-shirts and two movies.

Step 2. Look at [\[link\]](#). You can see from the fourth row/second column that three T-shirts are worth 63 utils. Similarly, the second row/fifth column shows that two movies are worth 31 utils.

Step 3. From this information, you can calculate that point Q has a total utility of 94 (63 + 31).

Step 4. You can repeat the same calculations for each point on [\[link\]](#), in which the total utility numbers are shown in the last column.

For José, the highest total utility for all possible combinations of goods occurs at point S, with a total utility of 103 from consuming one T-shirt and six movies.

## Choosing with Marginal Utility

Most people approach their utility-maximizing combination of choices in a step-by-step way. This step-by-step approach is based on looking at the tradeoffs, measured in terms of marginal utility, of consuming less of one good and more of another.

For example, say that José starts off thinking about spending all his money on T-shirts and choosing point P, which corresponds to four T-shirts and no movies, as illustrated in [\[link\]](#). José chooses this starting point randomly; he has to start somewhere. Then he considers giving up the last T-shirt, the one that provides him the least marginal utility, and using the money he saves to buy two movies instead. [\[link\]](#) tracks the step-by-step series of decisions José needs to make (Key: T-shirts are \$14, movies are \$7, and income is \$56). The following Work It Out feature explains how marginal utility can effect decision making.

Try	Which Has	Total Utility	Marginal Gain and Loss of Utility, Compared with Previous Choice	Conclusion
Choice 1: P	4 T-shirts and 0 movies	81 from 4 T-shirts + 0 from 0 movies = 81	–	–
Choice 2: Q	3 T-shirts and 2 movies	63 from 3 T-shirts + 31 from 0 movies = 94	Loss of 18 from 1 less T-shirt, but gain of 31 from 2 more movies, for a net utility gain of 13	Q is preferred over P
Choice 3: R	2 T-shirts and 4 movies	43 from 2 T-shirts + 58 from 4 movies = 101	Loss of 20 from 1 less T-shirt, but gain of 27 from two more movies for a net utility gain of 7	R is preferred over Q
Choice 4: S	1 T-shirt and 6 movies	22 from 1 T-shirt + 81 from 6 movies = 103	Loss of 21 from 1 less T-shirt, but gain of 23 from two more movies, for a net utility gain of 2	S is preferred over R
Choice 5: T	0 T-shirts and 8 movies	0 from 0 T-shirts + 100 from 8 movies = 100	Loss of 22 from 1 less T-shirt, but gain of 19 from two more movies, for a net utility loss of 3	S is preferred over T

## A Step-by-Step Approach to Maximizing Utility

### Note:

#### Decision Making by Comparing Marginal Utility

José could use the following thought process (if he thought in utils) to make his decision regarding how many T-shirts and movies to purchase:

Step 1. From [\[link\]](#), José can see that the marginal utility of the fourth T-shirt is 18. If José gives up the fourth T-shirt, then he loses 18 utils.

Step 2. Giving up the fourth T-shirt, however, frees up \$14 (the price of a T-shirt), allowing José to buy the first two movies (at \$7 each).

Step 3. José knows that the marginal utility of the first movie is 16 and the marginal utility of the second movie is 15. Thus, if José moves from point P to point Q, he gives up 18 utils (from the T-shirt), but gains 31 utils (from the movies).

Step 4. Gaining 31 utils and losing 18 utils is a net gain of 13. This is just another way of saying that the total utility at Q (94 according to the last column in [\[link\]](#)) is 13 more than the total utility at P (81).

Step 5. So, for José, it makes sense to give up the fourth T-shirt in order to buy two movies.

José clearly prefers point Q to point P. Now repeat this step-by-step process of decision making with marginal utilities. José thinks about giving up the third T-shirt and surrendering a marginal utility of 20, in exchange for purchasing two more movies that promise a combined marginal utility of 27. José prefers point R to point Q. What if José thinks about going beyond R to point S? Giving up the second T-shirt means a marginal utility loss of 21, and the marginal utility gain from the fifth and sixth movies would combine to make a marginal utility gain of 23, so José prefers point S to R.

However, if José seeks to go beyond point S to point T, he finds that the loss of marginal utility from giving up the first T-shirt is 22, while the marginal utility gain from the last two movies is only a total of 19. If José were to choose point T, his utility would fall to 100. Through these stages of thinking about marginal tradeoffs, José again concludes that S, with one T-shirt and six movies, is the choice that will provide him with the highest level of total utility. This step-by-step approach will reach the same conclusion regardless of José's starting point.

Another way to look at this is by focusing on satisfaction per dollar. **Marginal utility per dollar** is the amount of additional utility José receives given the price of the product. For José's T-shirts and movies, the marginal utility per dollar is shown in [\[link\]](#).

### Equation:

$$\text{marginal utility per dollar} = \frac{\text{marginal utility}}{\text{price}}$$

José's first purchase will be a movie. Why? Because it gives him the highest marginal utility per dollar and it is affordable. José will continue to purchase the good which gives him the highest marginal utility per dollar until he exhausts the budget. José will keep purchasing movies because they give him a greater "bang or the buck" until the sixth movie is equivalent to a T-shirt purchase. José can afford to purchase that T-shirt. So José will choose to purchase six movies and one T-shirt.

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
----------------------	---------------	------------------	-----------------------------	--------------------	---------------	------------------	-----------------------------

Quantity of T-Shirts	Total Utility	Marginal Utility	Marginal Utility per Dollar	Quantity of Movies	Total Utility	Marginal Utility	Marginal Utility per Dollar
1	22	22	22/\$14=1.6	1	16	16	16/\$7=2.3
2	43	21	21/\$14=1.5	2	31	15	15/\$7=2.1
3	63	20	20/\$14=1.4	3	45	14	14/\$7=2
4	81	18	18/\$14=1.3	4	58	13	13/\$7=1.9
5	97	16	16/\$14=1.1	5	70	12	12/\$7=1.7
6	111	14	14/\$14=1	6	81	11	11/\$7=1.6
7	123	12	12/\$14=1.2	7	91	10	10/\$7=1.4

Marginal Utility per Dollar

### A Rule for Maximizing Utility

This process of decision making suggests a rule to follow when maximizing utility. Since the price of T-shirts is twice as high as the price of movies, to maximize utility the last T-shirt chosen needs to provide exactly twice the marginal utility (MU) of the last movie. If the last T-shirt provides less than twice the marginal utility of the last movie, then the T-shirt is providing less “bang for the buck” (i.e., marginal utility per dollar spent) than if the same money were spent on movies. If this is so, José should trade the T-shirt for more movies to increase his total utility. Marginal utility per dollar measures the additional utility that José will enjoy given what he has to pay for the good.

If the last T-shirt provides more than twice the marginal utility of the last movie, then the T-shirt is providing more “bang for the buck” or marginal utility per dollar, than if the money were spent on movies. As a result, José should buy more T-shirts. Notice that at José’s optimal choice of point S, the marginal utility from the first T-shirt, of 22 is exactly twice the marginal utility of the sixth movie, which is 11. At this choice, the marginal utility per dollar is the same for both goods. This is a tell-tale signal that José has found the point with highest total utility.

This argument can be written as a general rule: the utility-maximizing choice between consumption goods occurs where the marginal utility per dollar is the same for both goods.

**Equation:**

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

A sensible economizer will pay twice as much for something only if, in the marginal comparison, the item confers twice as much utility. Notice that the formula for the table above is:

**Equation:**

$$\frac{22}{\$14} = \frac{11}{\$7}$$

$$1.6 = 1.6$$

The following Work It Out feature provides step by step guidance for this concept of utility-maximizing choices.

**Note:****Maximizing Utility**

The general rule,  $\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$ , means that the last dollar spent on each good provides exactly the same marginal utility. So:

Step 1. If we traded a dollar more of movies for a dollar more of T-shirts, the marginal utility gained from T-shirts would exactly offset the marginal utility lost from fewer movies. In other words, the net gain would be zero.

Step 2. Products, however, usually cost more than a dollar, so we cannot trade a dollar's worth of movies. The best we can do is trade two movies for another T-shirt, since in this example T-shirts cost twice what a movie does.

Step 3. If we trade two movies for one T-shirt, we would end up at point R (two T-shirts and four movies).

Step 4. Choice 4 in [\[link\]](#) shows that if we move to point S, we would lose 21 utils from one less T-shirt, but gain 23 utils from two more movies, so we would end up with more total utility at point S.

In short, the general rule shows us the utility-maximizing choice.

There is another, equivalent way to think about this. The general rule can also be expressed as *the ratio of the prices of the two goods should be equal to the ratio of the marginal utilities*. When the price of good 1 is divided by the price of good 2, at the utility-maximizing point this will equal the marginal utility of good 1 divided by the marginal utility of good 2. This rule, known as the **consumer equilibrium**, can be written in algebraic form:

**Equation:**

$$\frac{P_1}{P_2} = \frac{MU_1}{MU_2}$$

Along the budget constraint, the total price of the two goods remains the same, so the ratio of the prices does not change. However, the marginal utility of the two goods changes with the quantities consumed. At the optimal choice of one T-shirt and six movies, point S, the ratio of marginal utility to price for T-shirts (22:14) matches the ratio of marginal utility to price for movies (of 11:7).

**Measuring Utility with Numbers**

This discussion of utility started off with an assumption that it is possible to place numerical values on utility, an assumption that may seem questionable. You can buy a thermometer for measuring temperature at the hardware store, but what store sells an “utilimometer” for measuring utility? However, while measuring utility with numbers is a convenient assumption to clarify the explanation, the key assumption is not that utility can be measured by an outside party, but only that individuals can decide which of two alternatives they prefer.

To understand this point, think back to the step-by-step process of finding the choice with highest total utility by comparing the marginal utility that is gained and lost from different choices along the budget constraint. As José compares each choice along his budget constraint to the previous choice, what matters is not the specific numbers that he places on his utility—or whether he uses any numbers at all—but only that he personally can identify which choices he prefers.

In this way, the step-by-step process of choosing the highest level of utility resembles rather closely how many people make consumption decisions. We think about what will make us the happiest; we think about what things cost; we think about buying a little more of one item and giving up a little of something else; we choose what provides us with the greatest level of satisfaction. The vocabulary of comparing the points along a budget constraint and total and marginal utility is just a set of tools for discussing this everyday process in a clear and specific manner. It is welcome news that specific utility numbers are not central to the argument, since a good utilimometer is hard to find. Do not worry—while we cannot measure utils, by the end of the next module, we will have transformed our analysis into something we can measure—demand.

**Key Concepts and Summary**

Economic analysis of household behavior is based on the assumption that people seek the highest level of utility or satisfaction. Individuals are the only judge of their own utility. In general, greater consumption of a good brings higher total utility. However, the additional utility received from each unit of greater consumption tends to decline in a pattern of diminishing marginal utility.

The utility-maximizing choice on a consumption budget constraint can be found in several ways. You can add up total utility of each choice on the budget line and choose the highest total. You can choose a starting point at random and compare the marginal utility gains and losses of moving to neighboring points—and thus eventually seek out the preferred choice. Alternatively, you can compare the ratio of the marginal utility to price of good 1 with the marginal utility to price of good 2 and apply the rule that at the optimal choice, the two ratios should be equal:

**Equation:**

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$$

## Self-Check Questions

**Exercise:**

**Problem:**

Jeremy is deeply in love with Jasmine. Jasmine lives where cell phone coverage is poor, so he can either call her on the land-line phone for five cents per minute or he can drive to see her, at a round-trip cost of \$2 in gasoline money. He has a total of \$10 per week to spend on staying in touch. To make his preferred choice, Jeremy uses a handy utilimometer that measures his total utility from personal visits and from phone minutes. Using the values given in [\[link\]](#), figure out the points on Jeremy's consumption choice budget constraint (it may be helpful to do a sketch) and identify his utility-maximizing point.

Round Trips	Total Utility	Phone Minutes	Total Utility
0	0	0	0
1	80	20	200
2	150	40	380
3	210	60	540
4	260	80	680
5	300	100	800
6	330	120	900
7	200	140	980
8	180	160	1040
9	160	180	1080

Round Trips	Total Utility	Phone Minutes	Total Utility
10	140	200	1100

**Solution:**

The rows of the table in the problem do not represent the actual choices available on the budget set; that is, the combinations of round trips and phone minutes that Jeremy can afford with his budget. One of the choices listed in the problem, the six round trips, is not even available on the budget set. If Jeremy has only \$10 to spend and a round trip costs \$2 and phone calls cost \$0.05 per minute, he could spend his entire budget on five round trips but no phone calls or 200 minutes of phone calls, but no round trips or any combination of the two in between. It is easy to see all of his budget options with a little algebra. The equation for a budget line is:

**Equation:**

$$\text{Budget} = P_{\text{RT}} \times Q_{\text{RT}} + P_{\text{PC}} \times Q_{\text{PC}}$$

where P and Q are price and quantity of round trips ( $_{\text{RT}}$ ) and phone calls ( $_{\text{PC}}$ ) (per minute). In Jeremy's case the equation for the budget line is:

**Equation:**

$$\$10 = \$2 \times Q_{\text{RT}} + \$0.05 \times Q_{\text{PC}}$$

$$\frac{\$10}{\$0.05} = \frac{\$2Q_{\text{RT}} + \$0.05Q_{\text{PC}}}{\$0.05}$$

$$200 = 40Q_{\text{RT}} + Q_{\text{PC}}$$

$$Q_{\text{PC}} = 200 - 40Q_{\text{RT}}$$

If we choose zero through five round trips (column 1), the table below shows how many phone minutes can be afforded with the budget (column 3). The total utility figures are given in the table below.

Round Trips	Total Utility for Trips	Phone Minutes	Total Utility for Minutes	Total Utility
0	0	200	1100	1100
1	80	160	1040	1120
2	150	120	900	1050
3	210	80	680	890
4	260	40	380	640
5	300	0	0	300

Adding up total utility for round trips and phone minutes at different points on the budget line gives total utility at each point on the budget line. The highest possible utility is at the combination of one trip and 160

minutes of phone time, with a total utility of 1120.

**Exercise:**

**Problem:**

Take Jeremy's total utility information in [\[link\]](#), and use the marginal utility approach to confirm the choice of phone minutes and round trips that maximize Jeremy's utility.

**Solution:**

The first step is to use the total utility figures, shown in the table below, to calculate marginal utility, remembering that marginal utility is equal to the change in total utility divided by the change in trips or minutes.

Round Trips	Total Utility	Marginal Utility (per trip)	Phone Minutes	Total Utility	Marginal Utility (per minute)
0	0	-	200	1100	-
1	80	80	160	1040	$60/40 = 1.5$
2	150	70	120	900	$140/40 = 3.5$
3	210	60	80	680	$220/40 = 5.5$
4	260	50	40	380	$300/40 = 7.5$
5	300	40	0	0	$380/40 = 9.5$

Note that we cannot directly compare marginal utilities, since the units are trips versus phone minutes. We need a common denominator for comparison, which is price. Dividing MU by the price, yields columns 4 and 8 in the table below.

Round Trips	Total Utility	Marginal Utility (per trip)	MU/P	Phone Minutes	Total Utility	Marginal utility (per minute)	MU/P
0	0	-	-	200	1100	$60/40 = 1.5$	$1.5/\$0.05 = 30$
1	80	80	$80/\$2 = 40$	160	1040	$140/40 = 3.5$	$3.5/\$0.05 = 70$
2	150	70	$70/\$2 = 35$	120	900	$220/40 = 5.5$	$5.5/\$0.05 = 110$

Round Trips	Total Utility	Marginal Utility (per trip)	MU/P	Phone Minutes	Total Utility	Marginal utility (per minute)	MU/P
3	210	60	$60/\$2 = 30$	80	680	$300/40 = 7.5$	$7.5/\$0.05 = 150$
4	260	50	$50/\$2 = 25$	40	380	$380/40 = 9.5$	$9.5/\$0.05 = 190$
5	300	40	$40/\$2 = 20$	0	0	-	-

Start at the bottom of the table where the combination of round trips and phone minutes is (5, 0). This starting point is arbitrary, but the numbers in this example work best starting from the bottom. Suppose we consider moving to the next point up. At (4, 40), the marginal utility per dollar spent on a round trip is 25. The marginal utility per dollar spent on phone minutes is 190.

Since  $25 < 190$ , we are getting much more utility per dollar spent on phone minutes, so let's choose more of those. At (3, 80),  $MU/P_{RT}$  is  $30 < 150$  (the  $MU/P_M$ ), but notice that the difference is narrowing. We keep trading round trips for phone minutes until we get to (1, 160), which is the best we can do. The  $MU/P$  comparison is as close as it is going to get (40 vs. 70). Often in the real world, it is not possible to get  $MU/P$  exactly equal for both products, so you get as close as you can.

## Review Questions

### Exercise:

**Problem:** Who determines how much utility an individual will receive from consuming a good?

### Exercise:

**Problem:** Would you expect total utility to rise or fall with additional consumption of a good? Why?

### Exercise:

**Problem:** Would you expect marginal utility to rise or fall with additional consumption of a good? Why?

### Exercise:

**Problem:** Is it possible for total utility to increase while marginal utility diminishes? Explain.

### Exercise:

**Problem:**

If people do not have a complete mental picture of total utility for every level of consumption, how can they find their utility-maximizing consumption choice?

### Exercise:

**Problem:**

What is the rule relating the ratio of marginal utility to prices of two goods at the optimal choice? Explain why, if this rule does not hold, the choice cannot be utility-maximizing.

## Critical Thinking Questions

### Exercise:

#### Problem:

Think back to a purchase that you made recently. How would you describe your thinking before you made that purchase?

### Exercise:

#### Problem:

The rules of politics are not always the same as the rules of economics. In discussions of setting budgets for government agencies, there is a strategy called “closing the Washington monument.” When an agency faces the unwelcome prospect of a budget cut, it may decide to close a high-visibility attraction enjoyed by many people (like the Washington monument). Explain in terms of diminishing marginal utility why the Washington monument strategy is so misleading. *Hint:* If you are really trying to make the best of a budget cut, should you cut the items in your budget with the highest marginal utility or the lowest marginal utility? Does the Washington monument strategy cut the items with the highest marginal utility or the lowest marginal utility?

## Problems

### Exercise:

#### Problem:

Praxilla, who lived in ancient Greece, derives utility from reading poems and from eating cucumbers. Praxilla gets 30 units of marginal utility from her first poem, 27 units of marginal utility from her second poem, 24 units of marginal utility from her third poem, and so on, with marginal utility declining by three units for each additional poem. Praxilla gets six units of marginal utility for each of her first three cucumbers consumed, five units of marginal utility for each of her next three cucumbers consumed, four units of marginal utility for each of the following three cucumbers consumed, and so on, with marginal utility declining by one for every three cucumbers consumed. A poem costs three bronze coins but a cucumber costs only one bronze coin. Praxilla has 18 bronze coins. Sketch Praxilla’s budget set between poems and cucumbers, placing poems on the vertical axis and cucumbers on the horizontal axis. Start off with the choice of zero poems and 18 cucumbers, and calculate the changes in marginal utility of moving along the budget line to the next choice of one poem and 15 cucumbers. Using this step-by-step process based on marginal utility, create a table and identify Praxilla’s utility-maximizing choice. Compare the marginal utility of the two goods and the relative prices at the optimal choice to see if the expected relationship holds. *Hint:* Label the table columns: 1) Choice, 2) Marginal Gain from More Poems, 3) Marginal Loss from Fewer Cucumbers, 4) Overall Gain or Loss, 5) Is the previous choice optimal? Label the table rows: 1) 0 Poems and 18 Cucumbers, 2) 1 Poem and 15 Cucumbers, 3) 2 Poems and 12 Cucumbers, 4) 3 Poems and 9 Cucumbers, 5) 4 Poems and 6 Cucumbers, 6) 5 Poems and 3 Cucumbers, 7) 6 Poems and 0 Cucumbers.

## References

U.S. Bureau of Labor Statistics. 2015. “Consumer Expenditures in 2013.” Accessed March 11, 2015. <http://www.bls.gov/cex/csxann13.pdf>.

U.S. Bureau of Labor Statistics. 2015. “Employer Costs for Employee Compensation—December 2014.” Accessed March 11, 2015. <http://www.bls.gov/news.release/pdf/eccec.pdf>.

U.S. Bureau of Labor Statistics. 2015. “Labor Force Statistics from the Current Population Survey.” Accessed March 11, 2015. <http://www.bls.gov/cps/cpsaat22.htm>.

## Glossary

budget constraint line

shows the possible combinations of two goods that are affordable given a consumer's limited income

consumer equilibrium

when the ratio of the prices of goods is equal to the ratio of the marginal utilities (point at which the consumer can get the most satisfaction)

diminishing marginal utility

the common pattern that each marginal unit of a good consumed provides less of an addition to utility than the previous unit

marginal utility

the additional utility provided by one additional unit of consumption

marginal utility per dollar

the additional satisfaction gained from purchasing a good given the price of the product;  $MU/Price$

total utility

satisfaction derived from consumer choices

## How Changes in Income and Prices Affect Consumption Choices

By the end of this section, you will be able to:

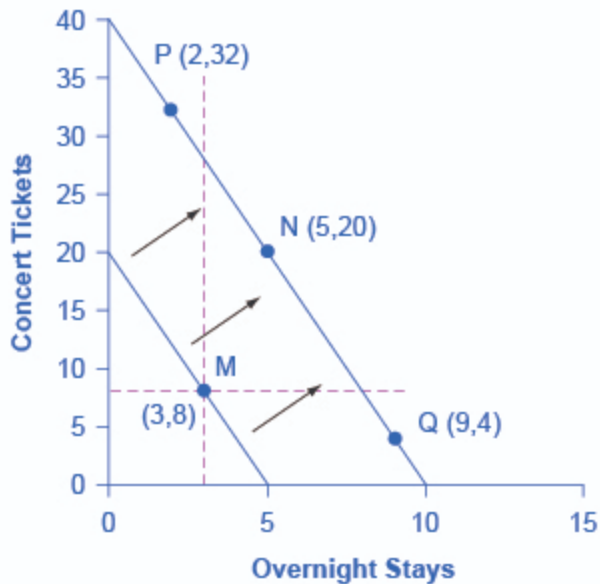
- Explain how income, prices, and preferences affect consumer choices
- Contrast the substitution effect and the income effect
- Utilize concepts of demand to analyze consumer choices
- Apply utility-maximizing choices to governments and businesses

Just as utility and marginal utility can be used to discuss making consumer choices along a budget constraint, these ideas can also be used to think about how consumer choices change when the budget constraint shifts in response to changes in income or price. Indeed, because the budget constraint framework can be used to analyze how quantities demanded change because of price movements, the budget constraint model can illustrate the underlying logic behind demand curves.

### How Changes in Income Affect Consumer Choices

Let's begin with a concrete example illustrating how changes in income level affect consumer choices. [\[link\]](#) shows a budget constraint that represents Kimberly's choice between concert tickets at \$50 each and getting away overnight to a bed-and-breakfast for \$200 per night. Kimberly has \$1,000 per year to spend between these two choices. After thinking about her total utility and marginal utility and applying the decision rule that the ratio of the marginal utilities to the prices should be equal between the two products, Kimberly chooses point M, with eight concerts and three overnight getaways as her utility-maximizing choice.

How a Change in Income Affects Consumption Choices



The utility-maximizing choice on the original budget constraint is M. The dashed horizontal and vertical lines extending through point M allow you to see at a glance whether the quantity consumed of goods on the new budget constraint is higher or lower than on the original budget constraint. On the new budget constraint, a choice like N will be made if both goods are normal goods. If overnight stays is an inferior good, a choice like P will be made. If concert tickets are an inferior good, a choice like Q will be made.

Now, assume that the income Kimberly has to spend on these two items rises to \$2,000 per year, causing her budget constraint to shift out to the right. How does this rise in income alter her utility-maximizing choice? Kimberly will again consider the utility and marginal utility that she receives from concert tickets and overnight getaways and seek her utility-maximizing choice on the new budget line. But how will her new choice relate to her original choice?

The possible choices along the new budget constraint can be divided into three groups, which are divided up by the dashed horizontal and vertical lines that pass through the original choice M in the figure. All choices on the upper left of the new budget constraint that are to the left of the vertical dashed line, like choice P with two overnight stays and 32 concert tickets, involve less of the good on the horizontal axis but much more of the good on the vertical axis. All choices to the right of the vertical dashed line and above the horizontal dashed line—like choice N with five overnight getaways and 20 concert tickets—have more consumption of both goods. Finally, all choices that are to the right of the vertical dashed line but below the horizontal dashed line, like choice Q with four concerts and nine overnight getaways, involve less of the good on the vertical axis but much more of the good on the horizontal axis.

All of these choices are theoretically possible, depending on Kimberly's personal preferences as expressed through the total and marginal utility she would receive from consuming these two goods. When income rises, the most common reaction is to purchase more of both goods, like choice N, which is to the upper right relative to Kimberly's original choice M, although exactly how much more of each good will vary according to personal taste. Conversely, when income falls, the most typical reaction is to purchase less of both goods. As defined in the chapter on [Demand and Supply](#) and again in the chapter on [Elasticity](#), goods and services are called normal goods when a rise in income leads to a rise in the quantity consumed of that good and a fall in income leads to a fall in quantity consumed.

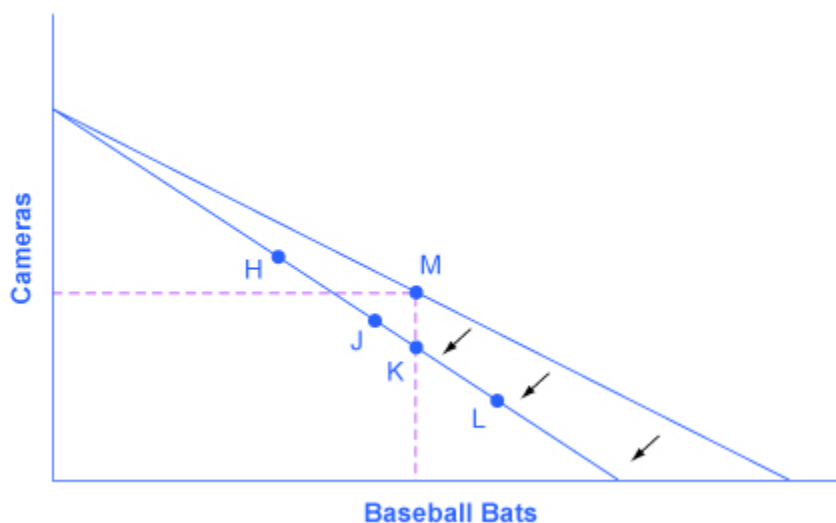
However, depending on Kimberly's preferences, a rise in income could cause consumption of one good to increase while consumption of the other good declines. A choice like P means that a rise in income caused her quantity consumed of overnight stays to decline, while a choice like Q would mean that a rise in income caused her quantity of concerts to decline. Goods where demand declines as income rises (or conversely, where the demand rises as income falls) are called "inferior goods." An inferior good occurs when people trim back on a good as income rises, because they can now afford the more expensive choices that they prefer. For example, a

higher-income household might eat fewer hamburgers or be less likely to buy a used car, and instead eat more steak and buy a new car.

## How Price Changes Affect Consumer Choices

For analyzing the possible effect of a change in price on consumption, let's again use a concrete example. [\[link\]](#) represents the consumer choice of Sergei, who chooses between purchasing baseball bats and cameras. A price increase for baseball bats would have no effect on the ability to purchase cameras, but it would reduce the number of bats Sergei could afford to buy. Thus a price increase for baseball bats, the good on the horizontal axis, causes the budget constraint to rotate inward, as if on a hinge, from the vertical axis. As in the previous section, the point labeled M represents the originally preferred point on the original budget constraint, which Sergei has chosen after contemplating his total utility and marginal utility and the tradeoffs involved along the budget constraint. In this example, the units along the horizontal and vertical axes are not numbered, so the discussion must focus on whether more or less of certain goods will be consumed, not on numerical amounts.

### How a Change in Price Affects Consumption Choices



The original utility-maximizing choice is M. When the price rises, the budget constraint shifts in to the left. The dashed lines make it possible to see at a glance whether the new consumption choice involves less of both goods, or less of one good and more of the

other. The new possible choices would be fewer baseball bats and more cameras, like point H, or less of both goods, as at point J. Choice K would mean that the higher price of bats led to exactly the same quantity of bats being consumed, but fewer cameras. Choices like L are ruled out as theoretically possible but highly unlikely in the real world, because they would mean that a higher price for baseball bats means a greater quantity consumed of baseball bats.

After the price increase, Sergei will make a choice along the new budget constraint. Again, his choices can be divided into three segments by the dashed vertical and horizontal lines. In the upper left portion of the new budget constraint, at a choice like H, Sergei consumes more cameras and fewer bats. In the central portion of the new budget constraint, at a choice like J, he consumes less of both goods. At the right-hand end, at a choice like L, he consumes more bats but fewer cameras.

The typical response to higher prices is that a person chooses to consume less of the product with the higher price. This occurs for two reasons, and both effects can occur simultaneously. The **substitution effect** occurs when a price changes and consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price. The **income effect** is that a higher price means, in effect, the buying power of income has been reduced (even though actual income has not changed), which leads to buying less of the good (when the good is normal). In this example, the higher price for baseball bats would cause Sergei to buy a fewer bats for both reasons. Exactly how much will a higher price for bats cause Sergei consumption of bats to fall? [\[link\]](#) suggests a range of possibilities. Sergei might react to a higher price for baseball bats by purchasing the same quantity of bats, but cutting his consumption of cameras. This choice is the point K on the new budget constraint, straight below the original choice M. Alternatively, Sergei might react by dramatically reducing his purchases of bats and instead buy more cameras.

The key is that it would be imprudent to assume that a change in the price of baseball bats will only or primarily affect the good whose price is changed, while the quantity consumed of other goods remains the same. Since Sergei purchases all his products out of the same budget, a change in the price of one good can also have a range of effects, either positive or negative, on the quantity consumed of other goods.

In short, a higher price typically causes reduced consumption of the good in question, but it can affect the consumption of other goods as well.

**Note:**

Read this [article](#) about the potential of variable prices in vending machines.

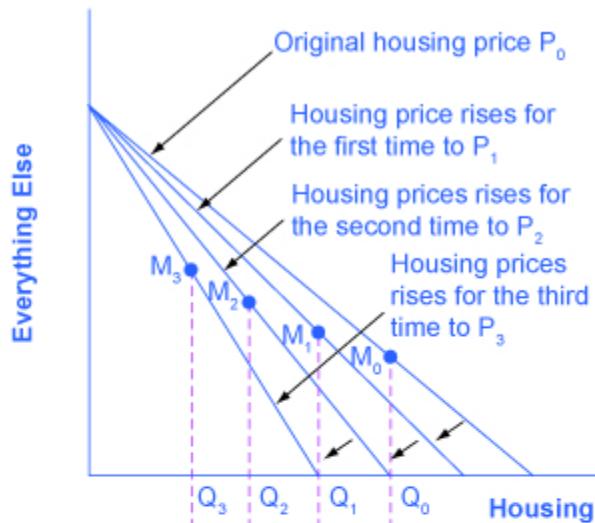


## The Foundations of Demand Curves

Changes in the price of a good lead the budget constraint to shift. A shift in the budget constraint means that when individuals are seeking their highest utility, the quantity that is demanded of that good will change. In this way, the logical foundations of demand curves—which show a connection between prices and quantity demanded—are based on the underlying idea of individuals seeking utility. [\[link\]](#) (a) shows a budget constraint with a choice between housing and “everything else.” (Putting “everything else” on the vertical axis can be a useful approach in some cases, especially when the focus of the analysis is on one particular good.) The preferred choice on the original budget constraint that provides the highest possible utility is labeled  $M_0$ . The other three budget constraints represent successively higher

prices for housing of  $P_1$ ,  $P_2$ , and  $P_3$ . As the budget constraint rotates in, and in, and in again, the utility-maximizing choices are labeled  $M_1$ ,  $M_2$ , and  $M_3$ , and the quantity demanded of housing falls from  $Q_0$  to  $Q_1$  to  $Q_2$  to  $Q_3$ .

### The Foundations of a Demand Curve: An Example of Housing



(a) A budget constraint diagram



(b) Deriving a demand curve

(a) As the price increases from  $P_0$  to  $P_1$  to  $P_2$  to  $P_3$ , the budget constraint on the upper part of the diagram shifts to the left. The utility-maximizing choice changes from  $M_0$  to  $M_1$  to  $M_2$  to  $M_3$ . As a result, the

quantity demanded of housing shifts from  $Q_0$  to  $Q_1$  to  $Q_2$  to  $Q_3$ , *ceteris paribus*. (b)

The demand curve graphs each combination of the price of housing and the quantity of housing demanded, *ceteris paribus*. Indeed, the quantities of housing are the same at the points on both (a) and (b). Thus, the original price of housing ( $P_0$ ) and the original quantity of housing ( $Q_0$ ) appear on the demand curve as point  $E_0$ .

The higher price of housing ( $P_1$ ) and the corresponding lower quantity demanded of housing ( $Q_1$ ) appear on the demand curve as point  $E_1$ .

So, as the price of housing rises, the budget constraint shifts to the left, and the quantity consumed of housing falls, *ceteris paribus* (meaning, with all other things being the same). This relationship—the price of housing rising from  $P_0$  to  $P_1$  to  $P_2$  to  $P_3$ , while the quantity of housing demanded falls from  $Q_0$  to  $Q_1$  to  $Q_2$  to  $Q_3$ —is graphed on the demand curve in [\[link\]](#) (b). Indeed, the vertical dashed lines stretching between the top and bottom of [\[link\]](#) show that the quantity of housing demanded at each point is the same in both (a) and (b). The shape of a demand curve is ultimately determined by the underlying choices about maximizing utility subject to a budget constraint. And while economists may not be able to measure “utils,” they can certainly measure price and quantity demanded.

## Applications in Government and Business

The budget constraint framework for making utility-maximizing choices offers a reminder that people can react to a change in price or income in a range of different ways. For example, in the winter months of 2005, costs for heating homes increased significantly in many parts of the country as prices for natural gas and electricity soared, due in large part to the disruption caused by Hurricanes Katrina and Rita. Some people reacted by

reducing the quantity demanded of energy; for example, by turning down the thermostats in their homes by a few degrees and wearing a heavier sweater inside. Even so, many home heating bills rose, so people adjusted their consumption in other ways, too. As you learned in the chapter on [Elasticity](#), the short run demand for home heating is generally inelastic. Each household cut back on what it valued least on the margin; for some it might have been some dinners out, or a vacation, or postponing buying a new refrigerator or a new car. Indeed, sharply higher energy prices can have effects beyond the energy market, leading to a widespread reduction in purchasing throughout the rest of the economy.

A similar issue arises when the government imposes taxes on certain products, like it does on gasoline, cigarettes, and alcohol. Say that a tax on alcohol leads to a higher price at the liquor store, the higher price of alcohol causes the budget constraint to pivot left, and consumption of alcoholic beverages is likely to decrease. However, people may also react to the higher price of alcoholic beverages by cutting back on other purchases. For example, they might cut back on snacks at restaurants like chicken wings and nachos. It would be unwise to assume that the liquor industry is the only one affected by the tax on alcoholic beverages. Read the next Clear It Up to learn about how buying decisions are influenced by who controls the household income.

**Note:**

**Does who controls household income make a difference?**

In the mid-1970s, the United Kingdom made an interesting policy change in its “child allowance” policy. This program provides a fixed amount of money per child to every family, regardless of family income.

Traditionally, the child allowance had been distributed to families by withholding less in taxes from the paycheck of the family wage earner—typically the father in this time period. The new policy instead provided the child allowance as a cash payment to the mother. As a result of this change, households have the same level of income and face the same prices in the market, but the money is more likely to be in the purse of the mother than in the wallet of the father.

Should this change in policy alter household consumption patterns? Basic models of consumption decisions, of the sort examined in this chapter, assume that it does not matter whether the mother or the father receives the money, because both parents seek to maximize the utility of the family as a whole. In effect, this model assumes that everyone in the family has the same preferences.

In reality, the share of income controlled by the father or the mother does affect what the household consumes. When the mother controls a larger share of family income a number of studies, in the United Kingdom and in a wide variety of other countries, have found that the family tends to spend more on restaurant meals, child care, and women's clothing, and less on alcohol and tobacco. As the mother controls a larger share of household resources, children's health improves, too. These findings suggest that when providing assistance to poor families, in high-income countries and low-income countries alike, the monetary amount of assistance is not all that matters: it also matters which member of the family actually receives the money.

The budget constraint framework serves as a constant reminder to think about the full range of effects that can arise from changes in income or price, not just effects on the one product that might seem most immediately affected.

## **Key Concepts and Summary**

The budget constraint framework suggest that when income or price changes, a range of responses are possible. When income rises, households will demand a higher quantity of normal goods, but a lower quantity of inferior goods. When the price of a good rises, households will typically demand less of that good—but whether they will demand a much lower quantity or only a slightly lower quantity will depend on personal preferences. Also, a higher price for one good can lead to more or less of the other good being demanded.

## **Self-Check Questions**

**Exercise:****Problem:**

Explain all the reasons why a decrease in the price of a product would lead to an increase in purchases of the product.

---

**Solution:**

This is the opposite of the example explained in the text. A decrease in price has a substitution effect and an income effect. The substitution effect says that because the product is cheaper relative to other things the consumer purchases, he or she will tend to buy more of the product (and less of the other things). The income effect says that after the price decline, the consumer could purchase the same goods as before, and still have money left over to purchase more. For both reasons, a decrease in price causes an increase in quantity demanded.

**Exercise:****Problem:**

As a college student you work at a part-time job, but your parents also send you a monthly “allowance.” Suppose one month your parents forgot to send the check. Show graphically how your budget constraint is affected. Assuming you only buy normal goods, what would happen to your purchases of goods?

---

**Solution:**

This is a negative income effect. Because your parents’ check failed to arrive, your monthly income is less than normal and your budget constraint shifts in toward the origin. If you only buy normal goods, the decrease in your income means you will buy less of every product.

**Review Questions****Exercise:**

**Problem:**

As a general rule, is it safe to assume that a change in the price of a good will always have its most significant impact on the quantity demanded of that good, rather than on the quantity demanded of other goods? Explain.

**Exercise:****Problem:**

Why does a change in income cause a parallel shift in the budget constraint?

**Critical Thinking Questions****Exercise:****Problem:**

Income effects depend on the income elasticity of demand for each good that you buy. If one of the goods you buy has a negative income elasticity, that is, it is an inferior good, what must be true of the income elasticity of the other good you buy?

**Problems****Exercise:****Problem:**

If a 10% decrease in the price of one product that you buy causes an 8% increase in quantity demanded of that product, will another 10% decrease in the price cause another 8% increase (no more and no less) in quantity demanded?

**Glossary**

income effect

a higher price means that, in effect, the buying power of income has been reduced, even though actual income has not changed; always happens simultaneously with a substitution effect

substitution effect

when a price changes, consumers have an incentive to consume less of the good with a relatively higher price and more of the good with a relatively lower price; always happens simultaneously with an income effect

## Labor-Leisure Choices

By the end of this section, you will be able to:

- Interpret labor-leisure budget constraint graphs
- Predict consumer choices based on wages and other compensation
- Explain the backward-bending supply curve of labor

People do not obtain utility just from products they purchase. They also obtain utility from leisure time. Leisure time is time not spent at work. The decision-making process of a utility-maximizing household applies to what quantity of hours to work in much the same way that it applies to purchases of goods and services. Choices made along the labor-leisure budget constraint, as wages shift, provide the logical underpinning for the labor supply curve. The discussion also offers some insights about the range of possible reactions when people receive higher wages, and specifically about the claim that if people are paid higher wages, they will work a greater quantity of hours—assuming that they have a say in the matter.

According to the Bureau of Labor Statistics, U.S. workers averaged 38.6 hours per week on the job in 2014. This average includes part-time workers; for full-time workers only, the average was 42.5 hours per week. [\[link\]](#) shows that more than half of all workers are on the job 35 to 48 hours per week, but significant proportions work more or less than this amount.

[\[link\]](#) breaks down the average hourly compensation received by private industry workers, including wages and benefits. Wages and salaries are about three-quarters of total compensation received by workers; the rest is in the form of health insurance, vacation pay, and other benefits. The compensation workers receive differs for many reasons, including experience, education, skill, talent, membership in a labor union, and the presence of discrimination against certain groups in the labor market. Issues surrounding the inequality of incomes in a market-oriented economy are explored in the chapters on [Poverty and Economic Inequality](#) and [Issues in Labor Markets: Unions, Discrimination, Immigration](#).

Hours Worked per Week	Number of Workers	Percentage of Workforce
1–14 hours	6.9 million	5.0%
15–34 hours	27.6 million	20.1%
35–40 hours	68.5 million	49.9%
41–48 hours	11.9 million	8.6%

Hours Worked per Week	Number of Workers	Percentage of Workforce
49–59 hours	13.3 million	9.6%
60 hours and over	9.3 million	6.8%

Persons at Work, by Average Hours Worked per Week in 2013 (Total number of workers: 137.7 million)(Source: <http://www.bls.gov/news.release/empsit.t18.htm>)

<b>Compensation, Wage, Salary, and Benefits</b>	<b>\$30.92 per hour</b>
<b>Wages and Salaries</b>	\$20.92
<b>Benefits</b>	
Vacation	\$2.09
Supplemental Pay	\$0.84
Insurance	\$2.15
Health Benefits	\$2.36
Retirement and Savings	\$1.24
Defined Benefit	\$0.57
Defined Contribution	\$0.064
Legally Required	\$2.46

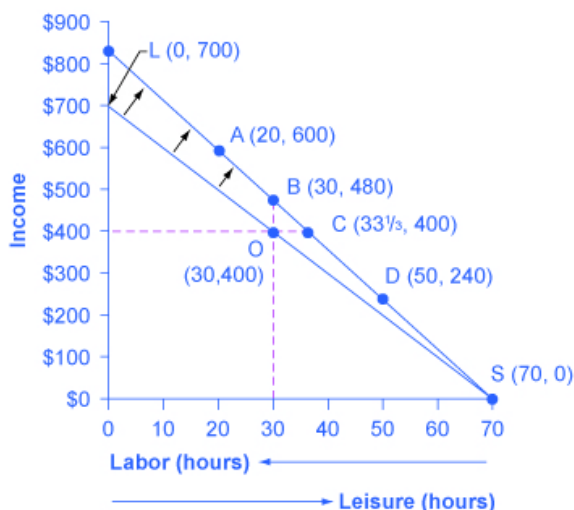
Hourly Compensation: Wages, Benefits, and Taxes in 2014(Source: <http://www.bls.gov/news.release/pdf/ecec.pdf>)

## The Labor-Leisure Budget Constraint

How do workers make decisions about the number of hours to work? Again, let's proceed with a concrete example. The economic logic is precisely the same as in the case of a consumption choice budget constraint, but the labels are different on a labor-leisure budget constraint.

Vivian has 70 hours per week that she could devote either to work or to leisure, and her wage is \$10/hour. The lower budget constraint in [\[link\]](#) shows Vivian's possible choices. The horizontal axis of this diagram measures both leisure and labor, by showing how Vivian's time is divided between leisure and labor. Hours of leisure are measured from left to right on the horizontal axis, while hours of labor are measured from right to left. Vivian will compare choices along this budget constraint, ranging from 70 hours of leisure and no income at point S to zero hours of leisure and \$700 of income at point L. She will choose the point that provides her with the highest total utility. For this example, let's assume that Vivian's utility-maximizing choice occurs at O, with 30 hours of leisure, 40 hours of work, and \$400 in weekly income.

#### How a Rise in Wages Alters the Utility-Maximizing Choice



Vivian's original choice is point O on the lower opportunity set. A rise in her wage causes her opportunity set to swing upward. In response to the increase in wages, Vivian can make a range of different choices available to her: a choice like D, which involves less work; and a choice like B, which involves the same amount of work but more income; or a choice like A, which involves more work and considerably more income. Vivian's personal preferences will determine which choice she makes.

For Vivian to discover the labor-leisure choice that will maximize her utility, she does not have to place numerical values on the total and marginal utility that she would receive from every level of income and leisure. All that really matters is that Vivian can compare, in her own mind, whether she would prefer more leisure or more income, given the tradeoffs she

faces. If Vivian can say to herself: “I’d really rather work a little less and have more leisure, even if it means less income,” or “I’d be willing to work more hours to make some extra income,” then as she gradually moves in the direction of her preferences, she will seek out the utility-maximizing choice on her labor-leisure budget constraint.

Now imagine that Vivian’s wage level increases to \$12/hour. A higher wage will mean a new budget constraint that tilts up more steeply; conversely, a lower wage would have led to a new budget constraint that was flatter. How will a change in the wage and the corresponding shift in the budget constraint affect Vivian’s decisions about how many hours to work?

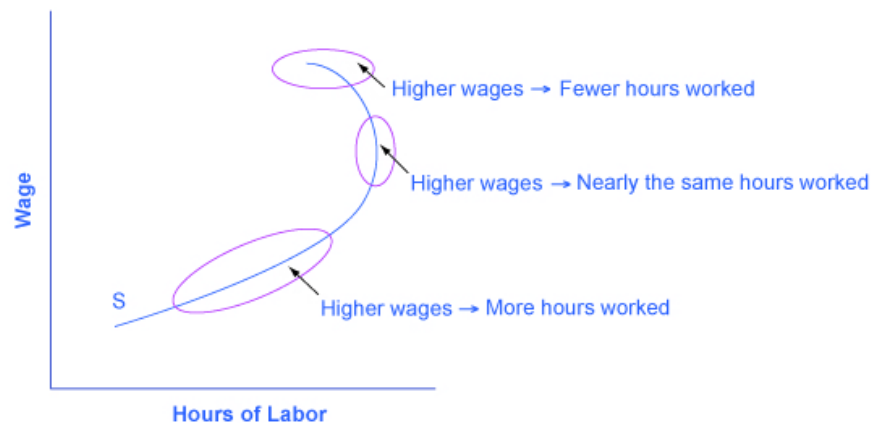
Vivian’s choices of quantity of hours to work and income along her new budget constraint can be divided into several categories, using the dashed horizontal and vertical lines in [\[link\]](#) that go through her original choice (O). One set of choices in the upper-left portion of the new budget constraint involves more hours of work (that is, less leisure) and more income, at a point like A with 20 hours of leisure, 50 hours of work, and \$600 of income (that is, 50 hours of work multiplied by the new wage of \$12 per hour). A second choice would be to work exactly the same 40 hours, and to take the benefits of the higher wage in the form of income that would now be \$480, at choice B. A third choice would involve more leisure and the same income at point C (that is, 33-1/3 hours of work multiplied by the new wage of \$12 per hour equals \$400 of total income). A fourth choice would involve less income and much more leisure at a point like D, with a choice like 50 hours of leisure, 20 hours of work, and \$240 in income.

In effect, Vivian can choose whether to receive the benefits of her wage increase in the form of more income, or more leisure, or some mixture of these two. With this range of possibilities, it would be unwise to assume that Vivian (or anyone else) will necessarily react to a wage increase by working substantially more hours. Maybe they will; maybe they will not.

## Applications of Utility Maximizing with the Labor-Leisure Budget Constraint

The theoretical insight that higher wages will sometimes cause an increase in hours worked, sometimes cause hours worked not to change by much, and sometimes cause hours worked to decline, has led to labor supply curves that look like the one in [\[link\]](#). The bottom-left portion of the labor supply curve slopes upward, which reflects the situation of a person who reacts to a higher wage by supplying a greater quantity of labor. The middle, close-to-vertical portion of the labor supply curve reflects the situation of a person who reacts to a higher wage by supplying about the same quantity of labor. The very top portion of the labor supply curve is called a **backward-bending supply curve for labor**, which is the situation of high-wage people who can earn so much that they respond to a still-higher wage by working fewer hours. Read the following Clear It Up feature for more on the number of hours the average person works each year.

### A Backward-Bending Supply Curve of Labor



The bottom upward-sloping portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked also increases. The middle, nearly vertical portion of the labor supply curve shows that as wages increase over this range, the quantity of hours worked changes very little. The backward-bending portion of the labor supply curve at the top shows that as wages increase over this range, the quantity of hours worked actually decreases. All three of these possibilities can be derived from how a change in wages causes movement in the labor-leisure budget constraint, and thus different choices by individuals.

#### **Note:**

##### **Is America a nation of workaholics?**

Americans work a lot. [\[link\]](#) shows average hours worked per year in the United States, Canada, Japan, and several European countries, with data from 2013. To get a perspective on these numbers, someone who works 40 hours per week for 50 weeks per year, with two weeks off, would work 2,000 hours per year. The gap in hours worked is a little astonishing; the 250 to 300 hour gap between how much Americans work and how much Germans or the French work amounts to roughly six to seven weeks less of work per year. Economists who study these international patterns debate the extent to which average Americans and Japanese have a preference for working more than, say, Germans, or whether German workers and employers face particular kinds of taxes and regulations that lead to fewer hours worked. Many countries have laws that regulate the work week and dictate holidays and the standards of “normal” vacation time vary from country to country. It is also interesting to take the amount of time spent working in context; it is estimated that in the late nineteenth century in the United States, the average work week was over 60 hours per week—leaving little to no time for leisure.

Country	Average Annual Hours Actually Worked per Employed Person
United States	1,824
Spain	1,799
Japan	1,759
Canada	1,751
United Kingdom	1,669
Sweden	1,585
Germany	1,443
France	1,441

Average Hours Worked Per Year in Select Countries(Source: <http://stats.oecd.org/Index.aspx?DataSetCode=ANHRS>)

The different responses to a rise in wages—more hours worked, the same hours worked, or fewer hours worked—are patterns exhibited by different groups of workers in the U.S. economy. Many full-time workers have jobs where the number of hours is held relatively fixed, partly by their own choice and partly by their employer's practices. These workers do not much change their hours worked as wages rise or fall, so their supply curve of labor is inelastic. However, part-time workers and younger workers tend to be more flexible in their hours, and more ready to increase hours worked when wages are high or cut back when wages fall.

The backward-bending supply curve for labor, when workers react to higher wages by working fewer hours and having more income, is not observed often in the short run. However, some well-paid professionals, like dentists or accountants, may react to higher wages by choosing to limit the number of hours, perhaps by taking especially long vacations, or taking every other Friday off. Over a long-term perspective, the backward-bending supply curve for labor is common. Over the last century, Americans have reacted to gradually rising wages by working fewer hours; for example, the length of the average work-week has fallen from about 60 hours per week in 1900 to the present average of less than 40 hours per week.

Recognizing that workers have a range of possible reactions to a change in wages casts some fresh insight on a perennial political debate: the claim that a reduction in income taxes—which

would, in effect, allow people to earn more per hour—will encourage people to work more. The leisure-income budget set points out that this connection will not hold true for all workers. Some people, especially part-timers, may react to higher wages by working more. Many will work the same number of hours. Some people, especially those whose incomes are already high, may react to the tax cut by working *fewer* hours. Of course, cutting taxes may be a good or a bad idea for a variety of reasons, not just because of its impact on work incentives, but the specific claim that tax cuts will lead people to work more hours is only likely to hold for specific groups of workers and will depend on how and for whom taxes are cut.

### Key Concepts and Summary

When making a choice along the labor-leisure budget constraint, a household will choose the combination of labor, leisure, and income that provides the most utility. The result of a change in wage levels can be higher work hours, the same work hours, or lower work hours.

### Self-Check Questions

**Exercise:**

**Problem:**

Siddhartha has 50 hours per week to devote to work or leisure. He has been working for \$8 per hour. Based on the information in [\[link\]](#), calculate his utility-maximizing choice of labor and leisure time.

Leisure Hours	Total Utility from Leisure	Work Hours	Income	Total Utility from Income
0	0	0	0	0
10	200	10	80	500
20	350	20	160	800
30	450	30	240	1,040
40	500	40	320	1,240
50	530	50	400	1,400

---

**Solution:**

This problem is straightforward if you remember leisure hours plus work hours are limited to 50 hours total. If you reverse the order of the last three columns so that more leisure corresponds to less work and income, you can add up columns two and five to find utility is maximized at 10 leisure hours and 40 work hours:

Leisure Hours	Total Utility from Leisure	Work Hours	Income	Total Utility from Income	Total Utility from Both
0	0	50	400	1,400	1,400
10	200	40	320	1,240	1,440
20	350	30	240	1,040	1,390
30	450	20	160	800	1,250
40	500	10	80	500	1,000
50	530	0	0	0	530

**Exercise:****Problem:**

In Siddhartha's problem, calculate marginal utility for income and for leisure. Now, start off at the choice with 50 hours of leisure and zero income, and a wage of \$8 per hour, and explain, in terms of marginal utility how Siddhartha could reason his way to the optimal choice, using marginal thinking only.

---

**Solution:**

Begin from the last table and compute marginal utility from leisure and work:

Leisure Hours	Total Utility from Leisure	MU from Leisure	Work Hours	Income	Total Utility from Income	MU from Income
0	0	-	50	400	1,400	160
10	200	200	40	320	1,240	200
20	350	150	30	240	1,040	240
30	450	100	20	160	800	300
40	500	50	10	80	500	500
50	530	30	0	0	0	-

Suppose Sid starts with 50 hours of leisure and 0 hours of work. As Sid moves up the table, he trades 10 hours of leisure for 10 hours of work at each step. At (40, 10), his  $MU_{\text{Leisure}} = 50$ , which is substantially less than his  $MU_{\text{Income}}$  of 500. This shortfall signals Sid to keep trading leisure for work/income until at (10, 40) the marginal utility of both is equal at 200. This is the sign that he should stop here, confirming the answer in question 1.

## Review Questions

### Exercise:

#### Problem:

How will a utility-maximizer find the choice of leisure and income that provides the greatest utility?

### Exercise:

#### Problem:

As a general rule, is it safe to assume that a higher wage will encourage significantly more hours worked for all individuals? Explain.

## Critical Thinking Questions

### Exercise:

**Problem:** In the labor-leisure choice model, what is the price of leisure?

**Exercise:****Problem:**

Think about the backward-bending part of the labor supply curve. Why would someone work less as a result of a higher wage rate?

**Exercise:****Problem:**

What would be the substitution effect and the income effect of a wage increase?

**Exercise:****Problem:**

Visit the BLS website and determine if education level, race/ethnicity, or gender appear to impact labor versus leisure choices.

**Glossary**

backward-bending supply curve for labor

the situation when high-wage people can earn so much that they respond to a still-higher wage by working fewer hours

## Intertemporal Choices in Financial Capital Markets

By the end of this section, you will be able to:

- Evaluate the reasons for making intertemporal choices
- Interpret an intertemporal budget constraint
- Analyze why people in America tend to save such a small percentage of their income

Rates of saving in America have never been especially high, but they seem to have dipped even lower in recent years, as the data from the Bureau of Economic Analysis in [\[link\]](#) show. A decision about how much to save can be represented using an intertemporal budget constraint. Household decisions about the quantity of financial savings show the same underlying pattern of logic as the consumption choice decision and the labor-leisure decision.

### Personal Savings as a Percentage of Personal Income



Personal savings were about 7 to 11% of personal income for most of the years from the late 1950s up to the early 1990s. Since then, the rate of personal savings has fallen substantially, although it seems to have bounced back a bit since 2008. (Source: <http://www.bea.gov/newsreleases/national/pi/pinewsrelease.htm>)

The discussion of financial saving here will not focus on the specific financial investment choices, like bank accounts, stocks, bonds, mutual funds, or owning a house or gold coins. The characteristics of these specific financial investments, along with the risks and tradeoffs they pose, are detailed in the [Labor and Financial Markets](#) chapter. Here, the focus is saving in total—that is, on how a household determines how much to consume in the present and how much to save, given the expected rate of return (or interest rate), and how the quantity of saving alters when the rate of return changes.

## Using Marginal Utility to Make Intertemporal Choices

Savings behavior varies considerably across households. One factor is that households with higher incomes tend to save a larger percentage of their income. This pattern makes intuitive sense; a well-to-do family has the flexibility in its budget to save 20–25% of income, while a poor family struggling to keep food on the table will find it harder to put money aside.

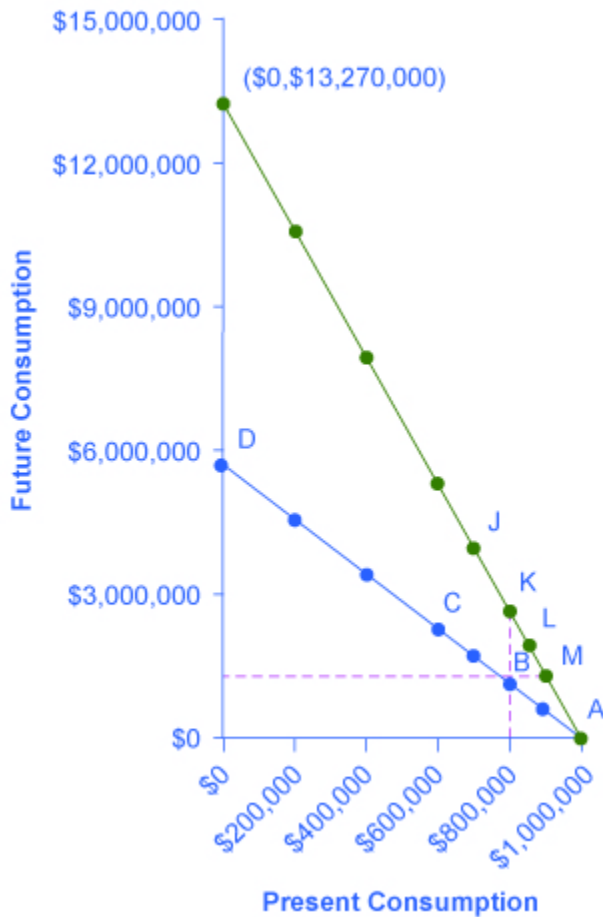
Another factor that causes personal saving to vary is personal preferences. Some people may prefer to consume more now, and let the future look after itself. Others may wish to enjoy a lavish retirement, complete with expensive vacations, or to pile up money that they can pass along to their grandchildren. There are savers and spendthrifts among the young, middle-aged, and old, and among those with high, middle, and low income levels.

Consider this example: Yelberton is a young man starting off at his first job. He thinks of the “present” as his working life and the “future” as after retirement. Yelberton’s plan is to save money from ages 30 to 60, retire at age 60, and then live off his retirement money from ages 60 to 85. On average, therefore, he will be saving for 30 years. If the rate of return that he can receive is 6% per year, then \$1 saved in the present would build up to \$5.74 after 30 years (using the formula for compound interest,  $\$1(1 + 0.06)^{30} = \$5.74$ ). Say that Yelberton will earn \$1,000,000 over the 30 years from age 30 to age 60 (this amount is approximately an annual salary of \$33,333 multiplied by 30 years). The question for Yelberton is how much of those lifetime earnings to consume during his working life, and how much to put aside until after retirement. This example is obviously built on

simplifying assumptions, but it does convey the basic life-cycle choice of saving during working life for future consumption after retirement.

[\[link\]](#) and [\[link\]](#) show Yelberton's intertemporal budget constraint. Yelberton's choice involves comparing the utility of present consumption during his working life and future consumption after retirement. The rate of return that determines the slope of the intertemporal budget line between present consumption and future consumption in this example is the annual interest rate that he would earn on his savings, compounded over the 30 years of his working life. (For simplicity, we are assuming that any savings from current income will compound for 30 years.) Thus, in the lower budget constraint line on the figure, future consumption grows by increments of \$574,000, because each time \$100,000 is saved in the present, it compounds to \$574,000 after 30 years at a 6% interest rate. If some of the numbers on the future consumption axis look bizarrely large, remember that this occurs because of the power of compound interest over substantial periods of time, and because the figure is grouping together all of Yelberton's saving for retirement over his lifetime.

Yelberton's Choice: The Intertemporal Budget Set



Yelberton will make a choice between present and future consumption. With an annual rate of return of 6%, he decides that his utility will be highest at point B, which represents a choice of \$800,000 in present consumption and \$1,148,000 in future consumption. When the annual rate of return rises to 9%, the intertemporal budget constraint pivots up. Yelberton could choose to take the gains from this higher rate of return in several forms: more present saving and much higher future consumption (J), the same present saving and higher future consumption (K), more present consumption and more future consumption (L), or more present

consumption and the same future consumption (M).

<b>Present Consumption</b>	<b>Present Savings</b>	<b>Future Consumption (6% annual return)</b>	<b>Future Consumption (9% annual return)</b>
\$1,000,000	0	0	0
\$900,000	\$100,000	\$574,000	\$1,327,000
\$800,000	\$200,000	\$1,148,000	\$2,654,000
\$700,000	\$300,000	\$1,722,000	\$3,981,000
\$600,000	\$400,000	\$2,296,000	\$5,308,000
\$400,000	\$600,000	\$3,444,000	\$7,962,000
\$200,000	\$800,000	\$4,592,000	\$10,616,000
0	\$1,000,000	\$5,740,000	\$13,270,000

### Yelburton's Intertemporal Budget Constraint

Yelberton will compare the different choices along the budget constraint and choose the one that provides him with the highest utility. For example, he will compare the utility he would receive from a choice like point A, with consumption of \$1 million in the present, zero savings, and zero future

consumption; point B, with present consumption of \$800,000, savings of \$200,000, and future consumption of \$1,148,000; point C, with present consumption of \$600,000, savings of \$400,000, and future consumption of \$2,296,000; or even choice D, with present consumption of zero, savings of \$1,000,000, and future consumption of \$5,740,000. Yelberton will also ask himself questions like these: “Would I prefer to consume a little less in the present, save more, and have more future consumption?” or “Would I prefer to consume a little more in the present, save less, and have less future consumption?” By considering marginal changes toward more or less consumption, he can seek out the choice that will provide him with the highest level of utility.

Let us say that Yelberton’s preferred choice is B. Imagine that Yelberton’s annual rate of return raises from 6% to 9%. In this case, each time he saves \$100,000 in the present, it will be worth \$1,327,000 in 30 years from now (using the formula for compound interest that  $\$100,000 (1 + 0.09)^{30} = \$1,327,000$ ). A change in rate of return alters the slope of the intertemporal budget constraint: a higher rate of return or interest rate will cause the budget line to pivot upward, while a lower rate of return will cause it to pivot downward. If Yelberton were to consume nothing in the present and save all \$1,000,000, with a 9% rate of return, his future consumption would be \$13,270,000, as shown on [\[link\]](#).

As the rate of return rises, Yelberton considers a range of choices on the new intertemporal budget constraint. The dashed vertical and horizontal lines running through the original choice B help to illustrate his range of options. One choice is to reduce present consumption (that is, to save more) and to have considerably higher future consumption at a point like J above and to the left of his original choice B. A second choice would be to keep the level of present consumption and savings the same, and to receive the benefits of the higher rate of return entirely in the form of higher future consumption, which would be choice K.

As a third choice Yelberton could have both more present consumption—that is, less savings—but still have higher future consumption because of the higher interest rate, which would be choice like L, above and to the right of his original choice B. Thus, the higher rate of return might cause

Yelberton to save more, or less, or the same amount, depending on his own preferences. A fourth choice would be that Yelberton could react to the higher rate of return by increasing his current consumption and leaving his future consumption unchanged, as at point M directly to the right of his original choice B. The actual choice of what quantity to save and how saving will respond to changes in the rate of return will vary from person to person, according to the choice that will maximize each person's utility.

## **Applications of the Model of Intertemporal Choice**

The theoretical model of the intertemporal budget constraint suggests that when the rate of return rises, the quantity of saving may rise, fall, or remain the same, depending on the preferences of individuals. For the U.S. economy as a whole, the most common pattern seems to be that the quantity of savings does not adjust much to changes in the rate of return. As a practical matter, many households either save at a fairly steady pace, by putting regular contributions into a retirement account or by making regular payments as they buy a house, or they do not save much at all. Of course, some people will have preferences that cause them to react to a higher rate of return by increasing their quantity of saving; others will react to a higher rate of return by noticing that with a higher rate of return, they can save less in the present and still have higher future consumption.

One prominent example in which a higher rate of return leads to a lower savings rate occurs when firms save money because they have promised to pay workers a certain fixed level of pension benefits after retirement. When rates of return rise, those companies can save less money in the present in their pension fund and still have enough to pay the promised retirement benefits in the future.

This insight suggests some skepticism about political proposals to encourage higher savings by providing savers with a higher rate of return. For example, Individual Retirement Accounts (IRAs) and 401(k) accounts are special savings accounts where the money going into the account is not taxed until it is taken out many years later, after retirement. The main difference between these accounts is that an IRA is usually set up by an individual, while a 401(k) needs to be set up through an employer. By not

taxing savings in the present, the effect of an IRA or a 401(k) is to increase the return to saving in these accounts.

IRA and 401(k) accounts have attracted a large quantity of savings since they became common in the late 1980s and early 1990s. In fact, the amount of IRAs rose from \$239 billion in 1992 to \$3.7 trillion in 2005, then to over \$5 trillion in 2012, as per the Investment Company Institute, a national association of U.S. investment companies. However, overall U.S. personal savings, as discussed earlier, actually dropped from low to lower in the late 1990s and into the 2000s. Evidently, the larger amounts in these retirement accounts are being offset, in the economy as a whole, either by less savings in other kinds of accounts, or by a larger amount of borrowing (that is, negative savings). The following Clear It Up further explores America's saving rates.

A rise in interest rates makes it easier for people to enjoy higher future consumption. But it also allows them to enjoy higher present consumption, if that is what these individuals desire. Again, a change in prices—in this case, in interest rates—leads to a range of possible outcomes.

**Note:**

How does America's saving rates compare to other countries?

By international standards, Americans do not save a high proportion of their income, as [\[link\]](#) shows. The rate of gross national saving includes saving by individuals, businesses, and government. By this measure, U.S. national savings amount to 17% of the size of the U.S. GDP, which measures the size of the U.S. economy. The comparable world average rate of savings is 22%.

<b>Country</b>	<b>Gross Domestic Savings as a Percentage of GDP</b>
China	51%
India	30%
Russia	28%
Mexico	22%
Germany	26%
Japan	22%
Canada	21%
France	21%
Brazil	15%
United States	17%
United Kingdom	13%
National Savings in Select Countries(Source: <a href="http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS">http://data.worldbank.org/indicator/NY.GNS.ICTR.ZS</a> )	

## **The Unifying Power of the Utility-Maximizing Budget Set Framework**

The choices of households are determined by an interaction between prices, budget constraints, and personal preferences. The flexible and powerful

terminology of utility-maximizing gives economists a vocabulary for bringing these elements together.

Not even economists believe that people walk around mumbling about their marginal utilities before they walk into a shopping mall, accept a job, or make a deposit in a savings account. However, economists do believe that individuals seek their own satisfaction or utility and that people often decide to try a little less of one thing and a little more of another. If these assumptions are accepted, then the idea of utility-maximizing households facing budget constraints becomes highly plausible.

## **Behavioral Economics: An Alternative Viewpoint**

As we know, people sometimes make decisions that seem “irrational” and not in their own best interest. People’s decisions can seem inconsistent from one day to the next and they even deliberately ignore ways to save money or time. The traditional economic models assume rationality, which means that people take all available information and make consistent and informed decisions that are in their best interest. (In fact, economics professors often delight in pointing out so-called “irrational behavior” each semester to their new students, and present economics as a way to become more rational.)

But a new group of economists, known as behavioral economists, argue that the traditional method leaves out something important: people’s state of mind. For example, one can think differently about money if one is feeling revenge, optimism, or loss. These are not necessarily irrational states of mind, but part of a range of emotions that can affect anyone on a given day. And what’s more, actions under these conditions are indeed predictable, if the underlying environment is better understood. So, **behavioral economics** seeks to enrich the understanding of decision-making by integrating the insights of psychology into economics. It does this by investigating how given dollar amounts can mean different things to individuals depending on the situation. This can lead to decisions that appear outwardly inconsistent, or irrational, to the outside observer.

The way the mind works, according to this view, may seem inconsistent to traditional economists but is actually far more complex than an unemotional

cost-benefit adding machine. For example, a traditional economist would say that if you lost a \$10 bill today, and also got an extra \$10 in your paycheck, you should feel perfectly neutral. After all,  $-\$10 + \$10 = \$0$ . You are the same financially as you were before. However, behavioral economists have done research that shows many people will feel some negative emotion—anger, frustration, and so forth—after those two things happen. We tend to focus more on the loss than the gain. This is known as loss aversion, where a \$1 loss pains us 2.25 times more than a \$1 gain helps us, according to the economists Daniel Kahneman and Amos Tversky in a famous 1979 article in the journal *Econometrica*. This insight has implications for investing, as people tend to “overplay” the stock market by reacting more to losses than to gains. Indeed, this behavior looks irrational to traditional economists, but is consistent once we understand better how the mind works, these economists argue.

Traditional economists also assume human beings have complete self-control. But, for instance, people will buy cigarettes by the pack instead of the carton even though the carton saves them money, to keep usage down. They purchase locks for their refrigerators and overpay on taxes to force themselves to save. In other words, we protect ourselves from our worst temptations but pay a price to do so. One way behavioral economists are responding to this is by setting up ways for people to keep themselves free of these temptations. This includes what are called “nudges” toward more rational behavior rather than mandatory regulations from government. For example, up to 20 percent of new employees do not enroll in retirement savings plans immediately, because of procrastination or feeling overwhelmed by the different choices. Some companies are now moving to a new system, where employees are automatically enrolled unless they “opt out.” Almost no-one opts out in this program and employees begin saving at the early years, which are most critical for retirement.

Another area that seems illogical is the idea of mental accounting, or putting dollars in different mental categories where they take different values. Economists typically consider dollars to be **fungible**, or having equal value to the individual, regardless of the situation.

You might, for instance, think of the \$25 you found in the street differently from the \$25 you earned from three hours working in a fast food restaurant. The street money might well be treated as “mad money” with little rational regard to getting the best value. This is in one sense strange, since it is still equivalent to three hours of hard work in the restaurant. Yet the “easy come-easy go” mentality replaces the rational economizer because of the situation, or context, in which the money was attained.

In another example of mental accounting that seems inconsistent to a traditional economist, a person could carry a credit card debt of \$1,000 that has a 15% yearly interest cost, and simultaneously have a \$2,000 savings account that pays only 2% per year. That means she pays \$150 a year to the credit card company, while collecting only \$40 annually in bank interest, so she loses \$130 a year. That doesn't seem wise.

The “rational” decision would be to pay off the debt, since a \$1,000 savings account with \$0 in debt is the equivalent net worth, and she would now net \$20 per year. But curiously, it is not uncommon for people to ignore this advice, since they will treat a loss to their savings account as higher than the benefit of paying off their credit card. The dollars are not being treated as fungible so it looks irrational to traditional economists.

Which view is right, the behavioral economists' or the traditional view? Both have their advantages, but behavioral economists have at least shed a light on trying to describe and explain behavior that has historically been dismissed as irrational. If most of us are engaged in some “irrational behavior,” perhaps there are deeper underlying reasons for this behavior in the first place.

**Note:**

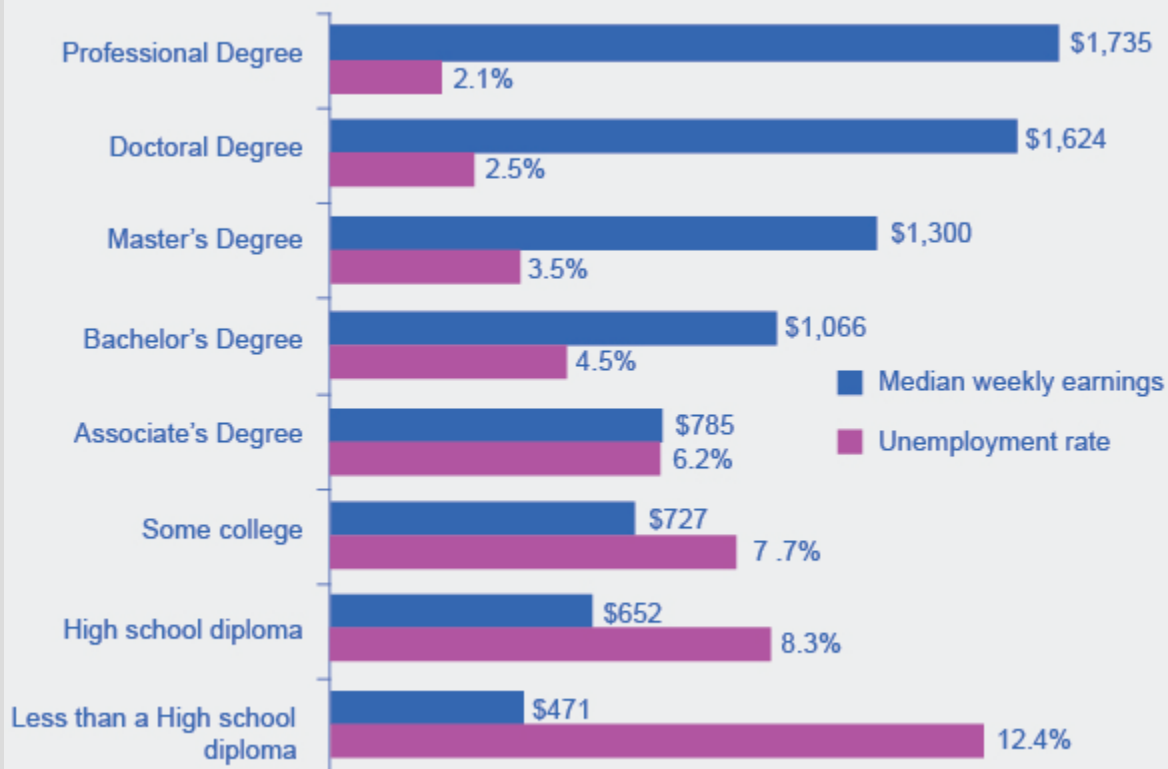
**"Eeny, Meeny, Miney, Moe"—Making Choices**

In what category did consumers worldwide increase their spending during the recession? Higher education. According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), enrollment in colleges and universities rose one-third in China and almost two-thirds in Saudi Arabia, nearly doubled in Pakistan, tripled in Uganda, and surged

by three million—18 percent—in the United States. Why were consumers willing to spend on education during lean times? Both individuals and countries view higher education as the way to prosperity. Many feel that increased earnings are a significant benefit of attending college.

Bureau of Labor Statistics data from May 2012 supports this view, as shown in [\[link\]](#). They show a positive correlation between earnings and education. The data also indicate that unemployment rates fall with higher levels of education and training.

#### The Impact of Education on Earnings and Unemployment Rates, 2012



Those with the highest degrees in 2012 had substantially lower unemployment rates whereas those with the least formal education suffered from the highest unemployment rates. The national median average weekly income was \$815, and the nation unemployment average in 2012 was 6.8%. (Source: Bureau of Labor Statistics, May 22, 2013)

## Key Concepts and Summary

When making a choice along the intertemporal budget constraint, a household will choose the combination of present consumption, savings, and future consumption that provides the most utility. The result of a higher rate of return (or higher interest rates) can be a higher quantity of saving, the same quantity of saving, or a lower quantity of saving, depending on preferences about present and future consumption. Behavioral economics is a branch of economics that seeks to understand and explain the "human" factors that drive what traditional economists see as people's irrational spending decisions.

## Self-Check Questions

### Exercise:

#### Problem:

How would an increase in expected income over one's lifetime affect one's intertemporal budget constraint? How would it affect one's consumption/saving decision?

---

#### Solution:

An increase in expected income would cause an outward shift in the intertemporal budget constraint. This would likely increase both current consumption and saving, but the answer would depend on one's time preference, that is, how much one is willing to wait to forgo current consumption. Children are notoriously bad at this, which is to say they might simply consume more, and not save any. Adults, because they think about the future, are generally better at time preference—that is, they are more willing to wait to receive a reward.

### Exercise:

**Problem:**

How would a decrease in expected interest rates over one's working life affect one's intertemporal budget constraint? How would it affect one's consumption/saving decision?

---

**Solution:**

Lower interest rates would make lending cheaper and saving less rewarding. This would be reflected in a flatter intertemporal budget line, a rotation around the amount of current income. This would likely cause a decrease in saving and an increase in current consumption, though the results for any individual would depend on time preference.

**Review Questions****Exercise:****Problem:**

According to the model of intertemporal choice, what are the major factors which determine how much saving an individual will do? What factors might a behavioral economist use to explain savings decisions?

**Exercise:****Problem:**

As a general rule, is it safe to assume that a lower interest rate will encourage significantly lower financial savings for all individuals? Explain.

**Critical Thinking Questions****Exercise:**

**Problem:**

What do you think accounts for the wide range of savings rates in different countries?

**Exercise:****Problem:**

What assumptions does the model of intertemporal choice make that are not likely true in the real world and would make the model harder to use in practice?

**References**

Holden, Sarah, and Daniel Schrass. 2012. "The rose of IRAs in U.S Households' Saving for Retirement, 2012." *ICI Research Perspective* 18.8 (2012). <http://www.ici.org/pdf/per18-08.pdf>.

Kahneman, Daniel and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47.2 (March 1979) 263-291.

Thaler, Richard H. "Shifting Our Retirement Savings into Automatic." *The New York Times*, April 6, 2013.  
<http://www.nytimes.com/2013/04/07/business/an-automatic-solution-for-the-retirement-savings-problem.html?pagewanted=all>.

UNESCO Institute for Statistics. "Statistics in Brief / Profiles" Accessed August 2013.  
[http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121 &IF\\_Language=en&BR\\_Country=5580](http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=121&IF_Language=en&BR_Country=5580).

**Glossary**

behavioral economics

a branch of economics that seeks to enrich the understanding of decision-making by integrating the insights of psychology and by

investigating how given dollar amounts can mean different things to individuals depending on the situation.

fungible

the idea that units of a good, such as dollars, ounces of gold, or barrels of oil are capable of mutual substitution with each other and carry equal value to the individual.

## Introduction to Cost and Industry Structure

class="introduction"

Amazon is an  
American  
international  
electronic  
commerce  
company that sells  
books, among  
many other things,  
shipping them  
directly to the  
consumer. There is  
no brick-and-  
mortar Amazon  
store. (Credit:  
modification of  
work by William  
Christiansen/Flick  
r Creative  
Commons)



**Note:**

**Amazon**

In less than two decades, Amazon.com has transformed the way books are sold, bought, and even read. Prior to Amazon, books were primarily sold through independent bookstores with limited inventories in small retail locations. There were exceptions, of course; Borders and Barnes & Noble offered larger stores in urban areas. In the last decade, however, independent bookstores have become few and far between, Borders has gone out of business, and Barnes & Noble is struggling. Online delivery and purchase of books has indeed overtaken the more traditional business models. How has Amazon changed the book selling industry? How has it managed to crush its competition?

A major reason for the giant retailer's success is its production model and cost structure, which has enabled Amazon to undercut the prices of its competitors even when factoring in the cost of shipping. Read on to see how firms great (like Amazon) and small (like your corner deli) determine what to sell, at what output and price.

**Note:**

## Introduction to Cost and Industry Structure

In this chapter, you will learn about:

- Explicit and Implicit Costs, and Accounting and Economic Profit
- The Structure of Costs in the Short Run
- The Structure of Costs in the Long Run

This chapter is the first of four chapters that explore the *theory of the firm*. This theory explains that firms behave in much the same way as consumers behave. What does that mean? Let's define what is meant by the firm. A **firm** (or business) combines inputs of labor, capital, land, and raw or finished component materials to produce outputs. If the firm is successful, the outputs are more valuable than the inputs. This activity of **production** goes beyond manufacturing (i.e., making things). It includes any process or service that creates value, including transportation, distribution, wholesale and retail sales. Production involves a number of important decisions that define the behavior of firms. These decisions include, but are not limited to:

- What product or products should the firm produce?
- How should the products be produced (i.e., what production process should be used)?
- How much output should the firm produce?
- What price should the firm charge for its products?
- How much labor should the firm employ?

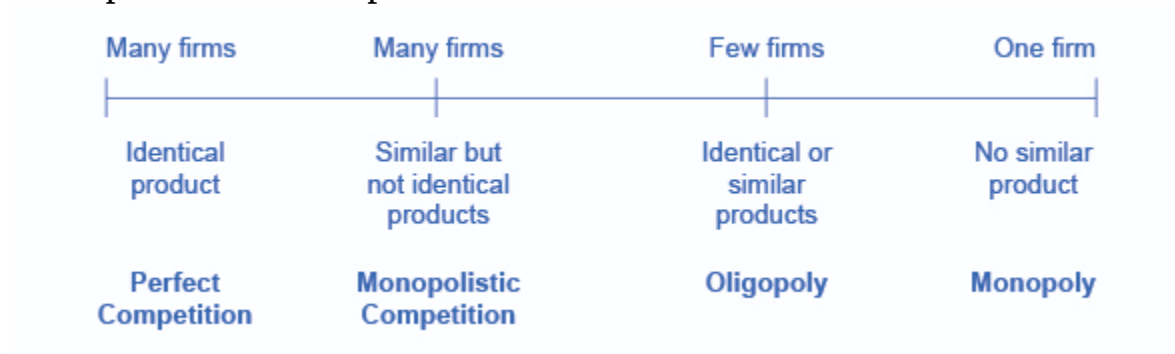
The answers to these questions depend on the production and cost conditions facing each firm. The answers also depend on the structure of the market for the product(s) in question. Market structure is a multidimensional concept that involves how competitive the industry is. It is defined by questions such as these:

- How much market power does each firm in the industry possess?
- How similar is each firm's product to the products of other firms in the industry?
- How difficult is it for new firms to enter the industry?

- Do firms compete on the basis of price, advertising, or other product differences?

[\[link\]](#) illustrates the range of different market structures, which we will explore in [Perfect Competition](#), [Monopoly](#), and [Monopolistic Competition and Oligopoly](#).

### The Spectrum of Competition



Firms face different competitive situations. At one extreme—perfect competition—many firms are all trying to sell identical products. At the other extreme—monopoly—only one firm is selling the product, and this firm faces no competition. Monopolistic competition and oligopoly fall between the extremes of perfect competition and monopoly. Monopolistic competition is a situation with many firms selling similar, but not identical, products. Oligopoly is a situation with few firms that sell identical or similar products.

First let's take a look at how firms determine their costs and desired profit levels. Then we will discuss costs in the short run and long run and the factors that can influence each.

## Explicit and Implicit Costs, and Accounting and Economic Profit

By the end of this section, you will be able to:

- Explain the difference between explicit costs and implicit costs
- Understand the relationship between cost and revenue

Private enterprise, the ownership of businesses by private individuals, is a hallmark of the U.S. economy. When people think of businesses, often giants like Wal-Mart, Microsoft, or General Motors come to mind. But firms come in all sizes, as shown in [\[link\]](#). The vast majority of American firms have fewer than 20 employees. As of 2010, the U.S. Census Bureau counted 5.7 million firms with employees in the U.S. economy. Slightly less than half of all the workers in private firms are at the 17,000 large firms, meaning they employ more than 500 workers. Another 35% of workers in the U.S. economy are at firms with fewer than 100 workers. These small-scale businesses include everything from dentists and lawyers to businesses that mow lawns or clean houses. Indeed, [\[link\]](#) does not include a separate category for the millions of small “non-employer” businesses where a single owner or a few partners are not officially paid wages or a salary, but simply receive whatever they can earn.

<b>Number of Employees</b>	<b>Firms (% of total firms)</b>	<b>Number of Paid Employees (% of total employment)</b>
<b>Total</b>	<b>5,734,538</b>	<b>112.0 million</b>
0–9	4,543,315 (79.2%)	12.3 million (11.0%)
10–19	617,089 (10.8%)	8.3 million (7.4%)

<b>Number of Employees</b>	<b>Firms (% of total firms)</b>	<b>Number of Paid Employees (% of total employment)</b>
20–99	475,125 (8.3%)	18.6 million (16.6%)
100–499	81,773 (1.4%)	15.9 million (14.2%)
500 or more	17,236 (0.30%)	50.9 million (49.8%)

Range in Size of U.S. Firms(Source: U.S. Census, 2010 [www.census.gov](http://www.census.gov))

Each of these businesses, regardless of size or complexity, tries to earn a profit:

**Equation:**

$$\text{Profit} = \text{Total Revenue} - \text{Total Cost}$$

Total **revenue** is the income brought into the firm from selling its products. It is calculated by multiplying the price of the product times the quantity of output sold:

**Equation:**

$$\text{Total Revenue} = \text{Price} \times \text{Quantity}$$

We will see in the following chapters that revenue is a function of the demand for the firm's products.

We can distinguish between two types of cost: explicit and implicit.

**Explicit costs** are out-of-pocket costs, that is, payments that are actually made. Wages that a firm pays its employees or rent that a firm pays for its office are explicit costs. **Implicit costs** are more subtle, but just as important. They represent the opportunity cost of using resources already owned by the firm. Often for small businesses, they are resources

contributed by the owners; for example, working in the business while not getting a formal salary, or using the ground floor of a home as a retail store. Implicit costs also allow for depreciation of goods, materials, and equipment that are necessary for a company to operate. (See the Work it Out feature for an extended example.)

These two definitions of cost are important for distinguishing between two conceptions of profit, accounting profit and economic profit. **Accounting profit** is a cash concept. It means total revenue minus explicit costs—the difference between dollars brought in and dollars paid out. **Economic profit** is total revenue minus total cost, including both explicit and implicit costs. The difference is important because even though a business pays income taxes based on its accounting profit, whether or not it is economically successful depends on its economic profit.

**Note:**

**Calculating Implicit Costs**

Consider the following example. Fred currently works for a corporate law firm. He is considering opening his own legal practice, where he expects to earn \$200,000 per year once he gets established. To run his own firm, he would need an office and a law clerk. He has found the perfect office, which rents for \$50,000 per year. A law clerk could be hired for \$35,000 per year. If these figures are accurate, would Fred's legal practice be profitable?

Step 1. First you have to calculate the costs. You can take what you know about explicit costs and total them:

**Equation:**

Office rental :	\$50,000
Law clerk's salary :	+\$35,000
	<hr/>
Total explicit costs :	\$85,000

Step 2. Subtracting the explicit costs from the revenue gives you the accounting profit.

**Equation:**

Revenues :	\$200,000
Explicit costs :	−\$85,000
Accounting profit :	\$115,000

But these calculations consider only the explicit costs. To open his own practice, Fred would have to quit his current job, where he is earning an annual salary of \$125,000. This would be an implicit cost of opening his own firm.

Step 3. You need to subtract both the explicit and implicit costs to determine the true economic profit:

**Equation:**

$$\begin{aligned}
 \text{Economic profit} &= \text{total revenues} - \text{explicit costs} - \text{implicit costs} \\
 &= \$200,000 - \$85,000 - \$125,000 \\
 &= -\$10,000 \text{ per year}
 \end{aligned}$$

Fred would be losing \$10,000 per year. That does not mean he would not want to open his own business, but it does mean he would be earning \$10,000 less than if he worked for the corporate firm.

Implicit costs can include other things as well. Maybe Fred values his leisure time, and starting his own firm would require him to put in more hours than at the corporate firm. In this case, the lost leisure would also be an implicit cost that would subtract from economic profits.

Now that we have an idea about the different types of costs, let's look at cost structures. A firm's cost structure in the long run may be different from that in the short run. We turn to that distinction in the next section.

## Key Concepts and Summary

Privately owned firms are motivated to earn profits. Profit is the difference between revenues and costs. While accounting profit considers only explicit costs, economic profit considers both explicit and implicit costs.

## Self-Check Questions

### Exercise:

#### Problem:

A firm had sales revenue of \$1 million last year. It spent \$600,000 on labor, \$150,000 on capital and \$200,000 on materials. What was the firm's accounting profit?

---

#### Solution:

Accounting profit = total revenues minus explicit costs = \$1,000,000 – (\$600,000 + \$150,000 + \$200,000) = \$50,000.

### Exercise:

#### Problem:

Continuing from [\[link\]](#), the firm's factory sits on land owned by the firm that could be rented out for \$30,000 per year. What was the firm's economic profit last year?

---

#### Solution:

Economic profit = accounting profit minus implicit cost = \$50,000 – \$30,000 = \$20,000.

## Review Questions

### Exercise:

**Problem:** What are explicit and implicit costs?

### Exercise:

**Problem:**

Would an interest payment on a loan to a firm be considered an explicit or implicit cost?

**Exercise:****Problem:**

What is the difference between accounting and economic profit?

**Critical Thinking Questions****Exercise:****Problem:**

Small “Mom and Pop firms,” like inner city grocery stores, sometimes exist even though they do not earn economic profits. How can you explain this?

**Problems****Exercise:****Problem:**

A firm is considering an investment that will earn a 6% rate of return. If it were to borrow the money, it would have to pay 8% interest on the loan, but it currently has the cash, so it will not need to borrow. Should the firm make the investment? Show your work.

**References**

2010 U.S. Census. [www.census.gov](http://www.census.gov).

**Glossary**

accounting profit

total revenues minus explicit costs, including depreciation

economic profit

total revenues minus total costs (explicit plus implicit costs)

explicit costs

out-of-pocket costs for a firm, for example, payments for wages and salaries, rent, or materials

firm

an organization that combines inputs of labor, capital, land, and raw or finished component materials to produce outputs.

implicit costs

opportunity cost of resources already owned by the firm and used in business, for example, expanding a factory onto land already owned

private enterprise

the ownership of businesses by private individuals

production

the process of combining inputs to produce outputs, ideally of a value greater than the value of the inputs

revenue

income from selling a firm's product; defined as price times quantity sold

## The Structure of Costs in the Short Run

By the end of this section, you will be able to:

- Analyze short-run costs as influenced by total cost, fixed cost, variable cost, marginal cost, and average cost.
- Calculate average profit
- Evaluate patterns of costs to determine potential profit

The cost of producing a firm's output depends on how much labor and physical capital the firm uses. A list of the costs involved in producing cars will look very different from the costs involved in producing computer software or haircuts or fast-food meals. However, the cost structure of all firms can be broken down into some common underlying patterns. When a firm looks at its **total costs** of production in the short run, a useful starting point is to divide total costs into two categories: fixed costs that cannot be changed in the short run and variable costs that can be changed.

### Fixed and Variable Costs

**Fixed costs** are expenditures that do not change regardless of the level of production, at least not in the short term. Whether you produce a lot or a little, the fixed costs are the same. One example is the rent on a factory or a retail space. Once you sign the lease, the rent is the same regardless of how much you produce, at least until the lease runs out. Fixed costs can take many other forms: for example, the cost of machinery or equipment to produce the product, research and development costs to develop new products, even an expense like advertising to popularize a brand name. The level of fixed costs varies according to the specific line of business: for instance, manufacturing computer chips requires an expensive factory, but a local moving and hauling business can get by with almost no fixed costs at all if it rents trucks by the day when needed.

**Variable costs**, on the other hand, are incurred in the act of producing—the more you produce, the greater the variable cost. Labor is treated as a variable cost, since producing a greater quantity of a good or service typically requires more workers or more work hours. Variable costs would also include raw materials.

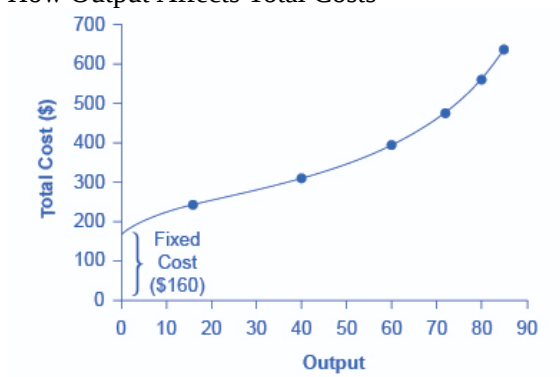
As a concrete example of fixed and variable costs, consider the barber shop called “The Clip Joint” shown in [\[link\]](#). The data for output and costs are shown in [\[link\]](#). The fixed costs of operating the barber shop, including the space and equipment, are \$160 per day. The variable costs are the costs of hiring barbers, which in our example is \$80 per barber each day. The first two columns of the table show the quantity of haircuts the barbershop can produce as it hires additional barbers. The third column shows the fixed costs, which do not change regardless of the level of production. The fourth column shows the variable costs at each level of output. These are calculated by taking the amount of labor hired and multiplying by the wage. For example, two barbers cost:  $2 \times \$80 = \$160$ . Adding together the fixed costs in the third column and the variable costs in the fourth column produces the total costs in the fifth column. So, for example, with two barbers the total cost is:  $\$160 + \$160 = \$320$ .

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost
1	16	\$160	\$80	\$240

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost
2	40	\$160	\$160	\$320
3	60	\$160	\$240	\$400
4	72	\$160	\$320	\$480
5	80	\$160	\$400	\$560
6	84	\$160	\$480	\$640
7	82	\$160	\$560	\$720

### Output and Total Costs

#### How Output Affects Total Costs



At zero production, the fixed costs of \$160 are still present. As production increases, variable costs are added to fixed costs, and the total cost is the sum of the two.

The relationship between the quantity of output being produced and the cost of producing that output is shown graphically in the figure. The fixed costs are always shown as the vertical intercept of the total cost curve; that is, they are the costs incurred when output is zero so there are no variable costs.

You can see from the graph that once production starts, total costs and variable costs rise. While variable costs may initially increase at a decreasing rate, at some point they begin increasing at an increasing rate. This is caused by diminishing marginal returns, discussed in the chapter on [Choice in a World of Scarcity](#), which is easiest to see with an example. As the number of barbers increases from zero to one in the table, output increases from 0 to 16 for a marginal gain of 16; as the number rises from one to two barbers, output increases from 16 to 40, a marginal gain of 24. From that point on, though, the marginal gain in output diminishes as each additional barber is added. For example, as the number of barbers rises from two to three, the marginal output gain is only 20; and as the number rises from three to four, the marginal gain is only 12.

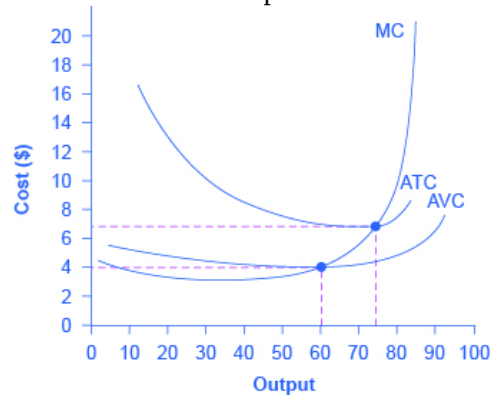
To understand the reason behind this pattern, consider that a one-man barber shop is a very busy operation. The single barber needs to do everything: say hello to people entering, answer the phone, cut hair, sweep up, and run the cash register. A second barber reduces the level of disruption from jumping back and forth between these tasks, and allows a greater division of labor and specialization. The result can be greater increasing marginal returns. However, as other barbers are added, the advantage of each additional barber is less, since the specialization of labor can only go so far. The addition of a sixth or seventh or eighth barber just to greet people at the door will have less impact than the second one did. This is the pattern of diminishing marginal returns. As a result, the total costs of production will begin to rise more rapidly as output increases. At some point, you may even see negative returns as the additional barbers begin bumping elbows and getting in each other's way. In this case, the addition of still more barbers would actually cause output to decrease, as shown in the last row of [\[link\]](#).

This pattern of diminishing marginal returns is common in production. As another example, consider the problem of irrigating a crop on a farmer's field. The plot of land is the fixed factor of production, while the water that can be added to the land is the key variable cost. As the farmer adds water to the land, output increases. But adding more and more water brings smaller and smaller increases in output, until at some point the water floods the field and actually reduces output. Diminishing marginal returns occur because, at a given level of fixed costs, each additional input contributes less and less to overall production.

### Average Total Cost, Average Variable Cost, Marginal Cost

The breakdown of total costs into fixed and variable costs can provide a basis for other insights as well. The first five columns of [\[link\]](#) duplicate the previous table, but the last three columns show average total costs, average variable costs, and marginal costs. These new measures analyze costs on a per-unit (rather than a total) basis and are reflected in the curves shown in [\[link\]](#).

Cost Curves at the Clip Joint



The information on total costs, fixed cost, and variable cost can also be presented on a per-unit basis.

Average total cost (ATC) is calculated by dividing total cost by the total quantity produced. The average total cost curve is typically U-shaped. Average variable cost (AVC) is calculated by dividing variable cost by the quantity produced. The average variable cost curve lies below the average total cost curve and is typically U-shaped

or upward-sloping. Marginal cost (MC) is calculated by taking the change in total cost between two levels of output and dividing by the change in output. The marginal cost curve is upward-sloping.

Labor	Quantity	Fixed Cost	Variable Cost	Total Cost	Marginal Cost	Average Total Cost	Average Variable Cost
1	16	\$160	\$80	\$240	\$5.00	\$15.00	\$5.00
2	40	\$160	\$160	\$320	\$3.30	\$8.00	\$4.00
3	60	\$160	\$240	\$400	\$4.00	\$6.60	\$4.00
4	72	\$160	\$320	\$480	\$6.60	\$6.60	\$4.40
5	80	\$160	\$400	\$560	\$10.00	\$7.00	\$5.00
6	84	\$160	\$480	\$640	\$20.00	\$7.60	\$5.70

### Different Types of Costs

**Average total cost** (sometimes referred to simply as average cost) is total cost divided by the quantity of output. Since the total cost of producing 40 haircuts is \$320, the average total cost for producing each of 40 haircuts is  $\$320/40$ , or \$8 per haircut. Average cost curves are typically U-shaped, as [link](#) shows. Average total cost starts off relatively high, because at low levels of output total costs are dominated by the fixed cost; mathematically, the denominator is so small that average total cost is large. Average total cost then declines, as the fixed costs are spread over an increasing quantity of output. In the average cost calculation, the rise in the numerator of total costs is relatively small compared to the rise in the denominator of quantity produced. But as output expands still further, the average cost begins to rise. At the right side of the average cost curve, total costs begin rising more rapidly as diminishing returns kick in.

**Average variable cost** obtained when variable cost is divided by quantity of output. For example, the variable cost of producing 80 haircuts is \$400, so the average variable cost is  $\$400/80$ , or \$5 per haircut. Note that at any level of output, the average variable cost curve will always lie below the curve for average total cost, as shown in [link](#). The reason is that average total cost includes average variable cost and average fixed cost. Thus, for  $Q = 80$  haircuts, the average total cost is \$8 per haircut, while the average variable cost is \$5 per haircut. However, as output grows, fixed costs become relatively less important (since they do not rise with output), so average variable cost sneaks closer to average cost.

Average total and variable costs measure the average costs of producing some quantity of output. Marginal cost is somewhat different. **Marginal cost** is the additional cost of producing one more unit of output. So it

is not the cost per unit of *all* units being produced, but only the next one (or next few). Marginal cost can be calculated by taking the change in total cost and dividing it by the change in quantity. For example, as quantity produced increases from 40 to 60 haircuts, total costs rise by  $400 - 320$ , or 80. Thus, the marginal cost for each of those marginal 20 units will be  $80/20$ , or \$4 per haircut. The marginal cost curve is generally upward-sloping, because diminishing marginal returns implies that additional units are more costly to produce. A small range of increasing marginal returns can be seen in the figure as a dip in the marginal cost curve before it starts rising. There is a point at which marginal and average costs meet, as the following Clear it Up feature discusses.

**Note:**

**Where do marginal and average costs meet?**

The marginal cost line intersects the average cost line exactly at the bottom of the average cost curve—which occurs at a quantity of 72 and cost of \$6.60 in [\[link\]](#). The reason why the intersection occurs at this point is built into the economic meaning of marginal and average costs. If the marginal cost of production is below the average cost for producing previous units, as it is for the points to the left of where MC crosses ATC, then producing one more additional unit will reduce average costs overall—and the ATC curve will be downward-sloping in this zone. Conversely, if the marginal cost of production for producing an additional unit is above the average cost for producing the earlier units, as it is for points to the right of where MC crosses ATC, then producing a marginal unit will increase average costs overall—and the ATC curve must be upward-sloping in this zone. The point of transition, between where MC is pulling ATC down and where it is pulling it up, must occur at the minimum point of the ATC curve.

This idea of the marginal cost “pulling down” the average cost or “pulling up” the average cost may sound abstract, but think about it in terms of your own grades. If the score on the most recent quiz you take is lower than your average score on previous quizzes, then the marginal quiz pulls down your average. If your score on the most recent quiz is higher than the average on previous quizzes, the marginal quiz pulls up your average. In this same way, low marginal costs of production first pull down average costs and then higher marginal costs pull them up.

The numerical calculations behind average cost, average variable cost, and marginal cost will change from firm to firm. However, the general patterns of these curves, and the relationships and economic intuition behind them, will not change.

## Lessons from Alternative Measures of Costs

Breaking down total costs into fixed cost, marginal cost, average total cost, and average variable cost is useful because each statistic offers its own insights for the firm.

Whatever the firm’s quantity of production, total revenue must exceed total costs if it is to earn a profit. As explored in the chapter [Choice in a World of Scarcity](#), fixed costs are often sunk costs that cannot be recouped. In thinking about what to do next, sunk costs should typically be ignored, since this spending has already been made and cannot be changed. However, variable costs can be changed, so they convey information about the firm’s ability to cut costs in the present and the extent to which costs will increase if production rises.

**Note:**

Why are total cost and average cost not on the same graph?

Total cost, fixed cost, and variable cost each reflect different aspects of the cost of production over the entire quantity of output being produced. These costs are measured in dollars. In contrast, marginal cost, average cost, and average variable cost are costs per unit. In the previous example, they are measured as cost per haircut. Thus, it would not make sense to put all of these numbers on the same graph, since they are measured in different units (\$ versus \$ per unit of output).

It would be as if the vertical axis measured two different things. In addition, as a practical matter, if they were on the same graph, the lines for marginal cost, average cost, and average variable cost would appear almost flat against the horizontal axis, compared to the values for total cost, fixed cost, and variable cost. Using the figures from the previous example, the total cost of producing 40 haircuts is \$320. But the average cost is \$320/40, or \$8. If you graphed both total and average cost on the same axes, the average cost would hardly show.

Average cost tells a firm whether it can earn profits given the current price in the market. If we divide profit by the quantity of output produced we get **average profit**, also known as the firm's *profit margin*. Expanding the equation for profit gives:

**Equation:**

$$\begin{aligned}\text{average profit} &= \frac{\text{profit}}{\text{quantity produced}} \\ &= \frac{\text{total revenue} - \text{total cost}}{\text{quantity produced}} \\ &= \frac{\text{total revenue}}{\text{quantity produced}} - \frac{\text{total cost}}{\text{quantity produced}} \\ &= \text{average revenue} - \text{average cost}\end{aligned}$$

But note that:

**Equation:**

$$\begin{aligned}\text{average revenue} &= \frac{\text{price} \times \text{quantity produced}}{\text{quantity produced}} \\ &= \text{price}\end{aligned}$$

Thus:

**Equation:**

$$\text{average profit} = \text{price} - \text{average cost}$$

This is the firm's profit margin. This definition implies that if the market price is above average cost, average profit, and thus total profit, will be positive; if price is below average cost, then profits will be negative.

The marginal cost of producing an additional unit can be compared with the marginal revenue gained by selling that additional unit to reveal whether the additional unit is adding to total profit—or not. Thus, marginal cost helps producers understand how profits would be affected by increasing or decreasing production.

## A Variety of Cost Patterns

The pattern of costs varies among industries and even among firms in the same industry. Some businesses have high fixed costs, but low marginal costs. Consider, for example, an Internet company that provides

medical advice to customers. Such a company might be paid by consumers directly, or perhaps hospitals or healthcare practices might subscribe on behalf of their patients. Setting up the website, collecting the information, writing the content, and buying or leasing the computer space to handle the web traffic are all fixed costs that must be undertaken before the site can work. However, when the website is up and running, it can provide a high quantity of service with relatively low variable costs, like the cost of monitoring the system and updating the information. In this case, the total cost curve might start at a high level, because of the high fixed costs, but then might appear close to flat, up to a large quantity of output, reflecting the low variable costs of operation. If the website is popular, however, a large rise in the number of visitors will overwhelm the website, and increasing output further could require a purchase of additional computer space.

For other firms, fixed costs may be relatively low. For example, consider firms that rake leaves in the fall or shovel snow off sidewalks and driveways in the winter. For fixed costs, such firms may need little more than a car to transport workers to homes of customers and some rakes and shovels. Still other firms may find that diminishing marginal returns set in quite sharply. If a manufacturing plant tried to run 24 hours a day, seven days a week, little time remains for routine maintenance of the equipment, and marginal costs can increase dramatically as the firm struggles to repair and replace overworked equipment.

Every firm can gain insight into its task of earning profits by dividing its total costs into fixed and variable costs, and then using these calculations as a basis for average total cost, average variable cost, and marginal cost. However, making a final decision about the profit-maximizing quantity to produce and the price to charge will require combining these perspectives on cost with an analysis of sales and revenue, which in turn requires looking at the market structure in which the firm finds itself. Before we turn to the analysis of market structure in other chapters, we will analyze the firm's cost structure from a long-run perspective.

## **Key Concepts and Summary**

In a short-run perspective, a firm's total costs can be divided into fixed costs, which a firm must incur before producing any output, and variable costs, which the firm incurs in the act of producing. Fixed costs are sunk costs; that is, because they are in the past and cannot be altered, they should play no role in economic decisions about future production or pricing. Variable costs typically show diminishing marginal returns, so that the marginal cost of producing higher levels of output rises.

Marginal cost is calculated by taking the change in total cost (or the change in variable cost, which will be the same thing) and dividing it by the change in output, for each possible change in output. Marginal costs are typically rising. A firm can compare marginal cost to the additional revenue it gains from selling another unit to find out whether its marginal unit is adding to profit.

Average total cost is calculated by taking total cost and dividing by total output at each different level of output. Average costs are typically U-shaped on a graph. If a firm's average cost of production is lower than the market price, a firm will be earning profits.

Average variable cost is calculated by taking variable cost and dividing by the total output at each level of output. Average variable costs are typically U-shaped. If a firm's average variable cost of production is lower than the market price, then the firm would be earning profits if fixed costs are left out of the picture.

## **Self-Check Questions**

### **Exercise:**

**Problem:**

The WipeOut Ski Company manufactures skis for beginners. Fixed costs are \$30. Fill in [\[link\]](#) for total cost, average variable cost, average total cost, and marginal cost.

Quantity	Variable Cost	Fixed Cost	Total Cost	Average Variable Cost	Average Total Cost	Marginal Cost
0	0	\$30				
1	\$10	\$30				
2	\$25	\$30				
3	\$45	\$30				
4	\$70	\$30				
5	\$100	\$30				
6	\$135	\$30				

---

**Solution:**

Quantity	Variable Cost	Fixed Cost	Total Cost	Average Variable Cost	Average Total Cost	Marginal Cost
0	0	\$30	\$30	-	-	
1	\$10	\$30	\$40	\$10.00	\$40.00	\$10
2	\$25	\$30	\$55	\$12.50	\$27.50	\$15
3	\$45	\$30	\$75	\$15.00	\$25.00	\$20
4	\$70	\$30	\$100	\$17.50	\$25.00	\$25
5	\$100	\$30	\$130	\$20.00	\$26.00	\$30

---

Quantity	Variable Cost	Fixed Cost	Total Cost	Average Variable Cost	Average Total Cost	Marginal Cost
6	\$135	\$30	\$165	\$22.50	\$27.50	\$35

**Exercise:**

**Problem:**

Based on your answers to the WipeOut Ski Company in [\[link\]](#), now imagine a situation where the firm produces a quantity of 5 units that it sells for a price of \$25 each.

- What will be the company's profits or losses?
- How can you tell at a glance whether the company is making or losing money at this price by looking at average cost?
- At the given quantity and price, is the marginal unit produced adding to profits?

---

**Solution:**

- Total revenues in this example will be a quantity of five units multiplied by the price of \$25/unit, which equals \$125. Total costs when producing five units are \$130. Thus, at this level of quantity and output the firm experiences losses (or negative profits) of \$5.
- If price is less than average cost, the firm is not making a profit. At an output of five units, the average cost is \$26/unit. Thus, at a glance you can see the firm is making losses. At a second glance, you can see that it must be losing \$1 for each unit produced (that is, average cost of \$26/unit minus the price of \$25/unit). With five units produced, this observation implies total losses of \$5.
- When producing five units, marginal costs are \$30/unit. Price is \$25/unit. Thus, the marginal unit is not adding to profits, but is actually subtracting from profits, which suggests that the firm should reduce its quantity produced.

## Review Questions

**Exercise:**

**Problem:** What is the difference between fixed costs and variable costs?

**Exercise:**

**Problem:** Are there fixed costs in the long-run? Explain briefly.

**Exercise:**

**Problem:** Are fixed costs also sunk costs? Explain.

**Exercise:**

**Problem:** What are diminishing marginal returns as they relate to costs?

**Exercise:****Problem:**

Which costs are measured on per-unit basis: fixed costs, average cost, average variable cost, variable costs, and marginal cost?

**Exercise:****Problem:**

How is each of the following calculated: marginal cost, average total cost, average variable cost?

**Critical Thinking Questions****Exercise:****Problem:**

A common name for fixed cost is “overhead.” If you divide fixed cost by the quantity of output produced, you get average fixed cost. Supposed fixed cost is \$1,000. What does the average fixed cost curve look like? Use your response to explain what “spreading the overhead” means.

**Exercise:**

**Problem:** How does fixed cost affect marginal cost? Why is this relationship important?

**Exercise:****Problem:**

Average cost curves (except for average fixed cost) tend to be U-shaped, decreasing and then increasing. Marginal cost curves have the same shape, though this may be harder to see since most of the marginal cost curve is increasing. Why do you think that average and marginal cost curves have the same general shape?

**Problems****Exercise:****Problem:**

Return to [\[link\]](#). What is the marginal gain in output from increasing the number of barbers from 4 to 5 and from 5 to 6? Does it continue the pattern of diminishing marginal returns?

**Exercise:****Problem:**

Compute the average total cost, average variable cost, and marginal cost of producing 60 and 72 haircuts. Draw the graph of the three curves between 60 and 72 haircuts.

**Glossary**

average profit

profit divided by the quantity of output produced; profit margin

average total cost

total cost divided by the quantity of output

average variable cost

variable cost divided by the quantity of output

fixed cost

expenditure that must be made before production starts and that does not change regardless of the level of production

marginal cost

the additional cost of producing one more unit

total cost

the sum of fixed and variable costs of production

variable cost

cost of production that increases with the quantity produced

## The Structure of Costs in the Long Run

By the end of this section, you will be able to:

- Calculate total cost
- Identify economies of scale, diseconomies of scale, and constant returns to scale
- Interpret graphs of long-run average cost curves and short-run average cost curves
- Analyze cost and production in the long run and short run

The long run is the period of time when all costs are variable. The long run depends on the specifics of the firm in question—it is not a precise period of time. If you have a one-year lease on your factory, then the long run is any period longer than a year, since after a year you are no longer bound by the lease. No costs are fixed in the long run. A firm can build new factories and purchase new machinery, or it can close existing facilities. In planning for the long run, the firm will compare alternative **production technologies** (or processes).

In this context, technology refers to all alternative methods of combining inputs to produce outputs. It does not refer to a specific new invention like the tablet computer. The firm will search for the production technology that allows it to produce the desired level of output at the lowest cost. After all, lower costs lead to higher profits—at least if total revenues remain unchanged. Moreover, each firm must fear that if it does not seek out the lowest-cost methods of production, then it may lose sales to competitor firms that find a way to produce and sell for less.

## Choice of Production Technology

Many tasks can be performed with a range of combinations of labor and physical capital. For example, a firm can have human beings answering phones and taking messages, or it can invest in an automated voicemail system. A firm can hire file clerks and secretaries to manage a system of paper folders and file cabinets, or it can invest in a computerized recordkeeping system that will require fewer employees. A firm can hire

workers to push supplies around a factory on rolling carts, it can invest in motorized vehicles, or it can invest in robots that carry materials without a driver. Firms often face a choice between buying a many small machines, which need a worker to run each one, or buying one larger and more expensive machine, which requires only one or two workers to operate it. In short, physical capital and labor can often substitute for each other.

Consider the example of a private firm that is hired by local governments to clean up public parks. Three different combinations of labor and physical capital for cleaning up a single average-sized park appear in [\[link\]](#). The first production technology is heavy on workers and light on machines, while the next two technologies substitute machines for workers. Since all three of these production methods produce the same thing—one cleaned-up park—a profit-seeking firm will choose the production technology that is least expensive, given the prices of labor and machines.

Production technology 1	10 workers	2 machines
Production technology 2	7 workers	4 machines
Production technology 3	3 workers	7 machines

Three Ways to Clean a Park

Production technology 1 uses the most labor and least machinery, while production technology 3 uses the least labor and the most machinery. [\[link\]](#) outlines three examples of how the total cost will change with each production technology as the cost of labor changes. As the cost of labor rises from example A to B to C, the firm will choose to substitute away from labor and use more machinery.

<b>Example A: Workers cost \$40, machines cost \$80</b>			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$40 = \$400$	$2 \times \$80 = \$160$	\$560
Cost of technology 2	$7 \times \$40 = \$280$	$4 \times \$80 = \$320$	\$600
Cost of technology 3	$3 \times \$40 = \$120$	$7 \times \$80 = \$560$	\$680
<b>Example B: Workers cost \$55, machines cost \$80</b>			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$55 = \$550$	$2 \times \$80 = \$160$	\$710
Cost of technology 2	$7 \times \$55 = \$385$	$4 \times \$80 = \$320$	\$705
Cost of technology 3	$3 \times \$55 = \$165$	$7 \times \$80 = \$560$	\$725
<b>Example C: Workers cost \$90, machines cost \$80</b>			
	Labor Cost	Machine Cost	Total Cost
Cost of technology 1	$10 \times \$90 = \$900$	$2 \times \$80 = \$160$	\$1,060
Cost of	$7 \times \$90 =$	$4 \times \$80 =$	\$950

technology 2	\$630	\$320	
Cost of technology 3	$3 \times \$90 =$ \$270	$7 \times \$80 =$ \$560	\$830

### Total Cost with Rising Labor Costs

Example A shows the firm's cost calculation when wages are \$40 and machines costs are \$80. In this case, technology 1 is the low-cost production technology. In example B, wages rise to \$55, while the cost of machines does not change, in which case technology 2 is the low-cost production technology. If wages keep rising up to \$90, while the cost of machines remains unchanged, then technology 3 clearly becomes the low-cost form of production, as shown in example C.

This example shows that as an input becomes more expensive (in this case, the labor input), firms will attempt to conserve on using that input and will instead shift to other inputs that are relatively less expensive. This pattern helps to explain why the demand curve for labor (or any input) slopes down; that is, as labor becomes relatively more expensive, profit-seeking firms will seek to substitute the use of other inputs. When a multinational employer like Coca-Cola or McDonald's sets up a bottling plant or a restaurant in a high-wage economy like the United States, Canada, Japan, or Western Europe, it is likely to use production technologies that conserve on the number of workers and focuses more on machines. However, that same employer is likely to use production technologies with more workers and less machinery when producing in a lower-wage country like Mexico, China, or South Africa.

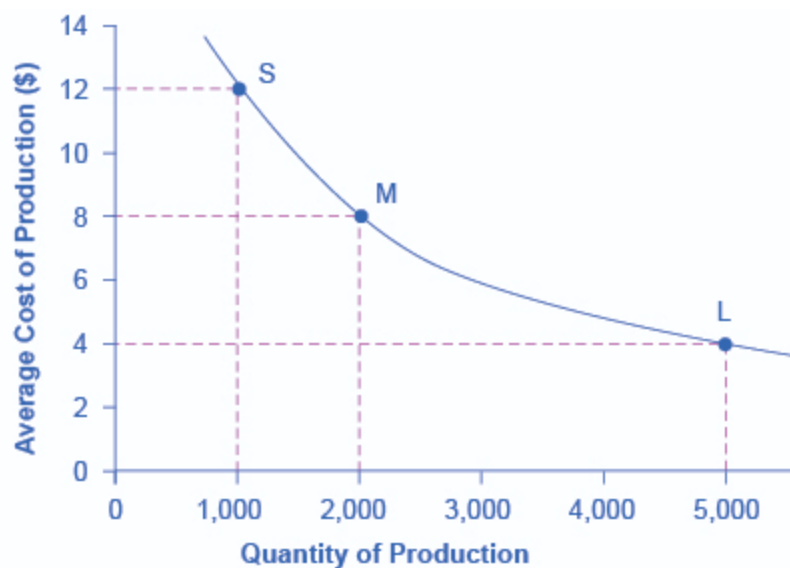
## Economies of Scale

Once a firm has determined the least costly production technology, it can consider the optimal scale of production, or quantity of output to produce. Many industries experience economies of scale. Economies of scale refers to the situation where, as the quantity of output goes up, the cost per unit goes down. This is the idea behind "warehouse stores" like Costco or

Walmart. In everyday language: a larger factory can produce at a lower average cost than a smaller factory.

[\[link\]](#) illustrates the idea of economies of scale, showing the average cost of producing an alarm clock falling as the quantity of output rises. For a small-sized factory like S, with an output level of 1,000, the average cost of production is \$12 per alarm clock. For a medium-sized factory like M, with an output level of 2,000, the average cost of production falls to \$8 per alarm clock. For a large factory like L, with an output of 5,000, the average cost of production declines still further to \$4 per alarm clock.

### Economies of Scale



A small factory like S produces 1,000 alarm clocks at an average cost of \$12 per clock. A medium factory like M produces 2,000 alarm clocks at a cost of \$8 per clock. A large factory like L produces 5,000 alarm clocks at a cost of \$4 per clock. Economies of scale exist because the larger scale of production leads to lower average costs.

The average cost curve in [\[link\]](#) may appear similar to the average cost curves presented earlier in this chapter, although it is downward-sloping rather than U-shaped. But there is one major difference. The economies of scale curve is a long-run average cost curve, because it allows all factors of

production to change. The short-run average cost curves presented earlier in this chapter assumed the existence of fixed costs, and only variable costs were allowed to change.

One prominent example of economies of scale occurs in the chemical industry. Chemical plants have a lot of pipes. The cost of the materials for producing a pipe is related to the circumference of the pipe and its length. However, the volume of chemicals that can flow through a pipe is determined by the cross-section area of the pipe. The calculations in [\[link\]](#) show that a pipe which uses twice as much material to make (as shown by the circumference of the pipe doubling) can actually carry four times the volume of chemicals because the cross-section area of the pipe rises by a factor of four (as shown in the Area column).

	Circumference ( $2\pi r$ )	Area ( $\pi r^2$ )
4-inch pipe	12.5 inches	12.5 square inches
8-inch pipe	25.1 inches	50.2 square inches
16-inch pipe	50.2 inches	201.1 square inches

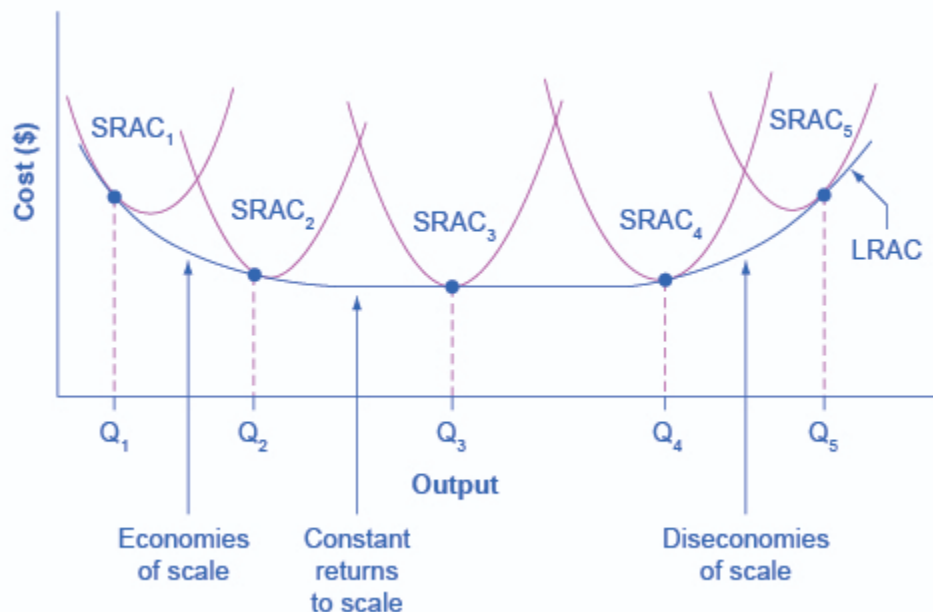
Comparing Pipes: Economies of Scale in the Chemical Industry

A doubling of the cost of producing the pipe allows the chemical firm to process four times as much material. This pattern is a major reason for economies of scale in chemical production, which uses a large quantity of pipes. Of course, economies of scale in a chemical plant are more complex than this simple calculation suggests. But the chemical engineers who design these plants have long used what they call the “six-tenths rule,” a rule of thumb which holds that increasing the quantity produced in a chemical plant by a certain percentage will increase total cost by only six-tenths as much.

## Shapes of Long-Run Average Cost Curves

While in the short run firms are limited to operating on a single average cost curve (corresponding to the level of fixed costs they have chosen), in the long run when all costs are variable, they can choose to operate on any average cost curve. Thus, the **long-run average cost (LRAC) curve** is actually based on a group of **short-run average cost (SRAC) curves**, each of which represents one specific level of fixed costs. More precisely, the long-run average cost curve will be the least expensive average cost curve for any level of output. [\[link\]](#) shows how the long-run average cost curve is built from a group of short-run average cost curves. Five short-run-average cost curves appear on the diagram. Each SRAC curve represents a different level of fixed costs. For example, you can imagine  $SRAC_1$  as a small factory,  $SRAC_2$  as a medium factory,  $SRAC_3$  as a large factory, and  $SRAC_4$  and  $SRAC_5$  as very large and ultra-large. Although this diagram shows only five SRAC curves, presumably there are an infinite number of other SRAC curves between the ones that are shown. This family of short-run average cost curves can be thought of as representing different choices for a firm that is planning its level of investment in fixed cost physical capital—knowing that different choices about capital investment in the present will cause it to end up with different short-run average cost curves in the future.

From Short-Run Average Cost Curves to Long-Run Average Cost Curves



The five different short-run average cost (SRAC) curves each represents a different level of fixed costs, from the low level of fixed costs at  $SRAC_1$  to the high level of fixed costs at  $SRAC_5$ . Other SRAC curves, not shown in the diagram, lie between the ones that are shown here. The long-run average cost (LRAC) curve shows the lowest cost for producing each quantity of output when fixed costs can vary, and so it is formed by the bottom edge of the family of SRAC curves. If a firm wished to produce quantity  $Q_3$ , it would choose the fixed costs associated with  $SRAC_3$ .

The long-run average cost curve shows the cost of producing each quantity in the long run, when the firm can choose its level of fixed costs and thus choose which short-run average costs it desires. If the firm plans to produce in the long run at an output of  $Q_3$ , it should make the set of investments that will lead it to locate on  $SRAC_3$ , which allows producing  $q_3$  at the lowest cost. A firm that intends to produce  $Q_3$  would be foolish to choose the level of fixed costs at  $SRAC_2$  or  $SRAC_4$ . At  $SRAC_2$  the level of fixed costs is too low for producing  $Q_3$  at lowest possible cost, and producing  $q_3$  would require adding a very high level of variable costs and make the average cost very high. At  $SRAC_4$ , the level of fixed costs is too high for producing  $q_3$  at lowest possible cost, and again average costs would be very high as a result.

The shape of the long-run cost curve, as drawn in [\[link\]](#), is fairly common for many industries. The left-hand portion of the long-run average cost curve, where it is downward-sloping from output levels  $Q_1$  to  $Q_2$  to  $Q_3$ , illustrates the case of economies of scale. In this portion of the long-run average cost curve, larger scale leads to lower average costs. This pattern was illustrated earlier in [\[link\]](#).

In the middle portion of the long-run average cost curve, the flat portion of the curve around  $Q_3$ , economies of scale have been exhausted. In this situation, allowing all inputs to expand does not much change the average cost of production, and it is called **constant returns to scale**. In this range of the LRAC curve, the average cost of production does not change much

as scale rises or falls. The following Clear it Up feature explains where diminishing marginal returns fit into this analysis.

**Note:**

How do economies of scale compare to diminishing marginal returns?

The concept of economies of scale, where average costs decline as production expands, might seem to conflict with the idea of diminishing marginal returns, where marginal costs rise as production expands. But diminishing marginal returns refers only to the short-run average cost curve, where one variable input (like labor) is increasing, but other inputs (like capital) are fixed. Economies of scale refers to the long-run average cost curve where all inputs are being allowed to increase together. Thus, it is quite possible and common to have an industry that has both diminishing marginal returns when only one input is allowed to change, and at the same time has increasing or constant economies of scale when all inputs change together to produce a larger-scale operation.

Finally, the right-hand portion of the long-run average cost curve, running from output level  $Q_4$  to  $Q_5$ , shows a situation where, as the level of output and the scale rises, average costs rise as well. This situation is called **diseconomies of scale**. A firm or a factory can grow so large that it becomes very difficult to manage, resulting in unnecessarily high costs as many layers of management try to communicate with workers and with each other, and as failures to communicate lead to disruptions in the flow of work and materials. Not many overly large factories exist in the real world, because with their very high production costs, they are unable to compete for long against plants with lower average costs of production. However, in some planned economies, like the economy of the old Soviet Union, plants that were so large as to be grossly inefficient were able to continue operating for a long time because government economic planners protected them from competition and ensured that they would not make losses.

Diseconomies of scale can also be present across an entire firm, not just a large factory. The leviathan effect can hit firms that become too large to run

efficiently, across the entirety of the enterprise. Firms that shrink their operations are often responding to finding itself in the diseconomies region, thus moving back to a lower average cost at a lower output level.

**Note:**

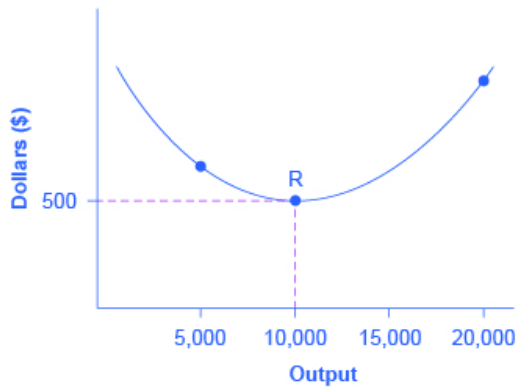
Visit this [website](#) to read an article about the complexity of the belief that banks can be “too-big-to-fail.”



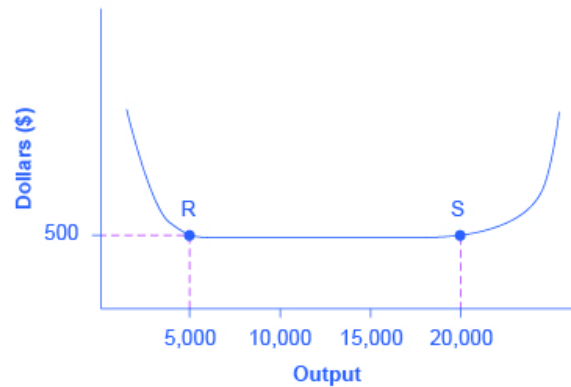
## The Size and Number of Firms in an Industry

The shape of the long-run average cost curve has implications for how many firms will compete in an industry, and whether the firms in an industry have many different sizes, or tend to be the same size. For example, say that one million dishwashers are sold every year at a price of \$500 each and the long-run average cost curve for dishwashers is shown in [\[link\]](#) (a). In [\[link\]](#) (a), the lowest point of the LRAC curve occurs at a quantity of 10,000 produced. Thus, the market for dishwashers will consist of 100 different manufacturing plants of this same size. If some firms built a plant that produced 5,000 dishwashers per year or 25,000 dishwashers per year, the average costs of production at such plants would be well above \$500, and the firms would not be able to compete.

The LRAC Curve and the Size and Number of Firms



(a) LRAC curve with a clear minimum point



(b) A flat-bottomed LRAC curve

(a) Low-cost firms will produce at output level R. When the LRAC curve has a clear minimum point, then any firm producing a different quantity will have higher costs. In this case, a firm producing at a quantity of 10,000 will produce at a lower average cost than a firm producing, say, 5,000 or 20,000 units. (b) Low-cost firms will produce between output levels R and S. When the LRAC curve has a flat bottom, then firms producing at any quantity along this flat bottom can compete. In this case, any firm producing a quantity between 5,000 and 20,000 can compete effectively, although firms producing less than 5,000 or more than 20,000 would face higher average costs and be unable to compete.

**Note:**

How can cities be viewed as examples of economies of scale?

Why are people and economic activity concentrated in cities, rather than distributed evenly across a country? The fundamental reason must be related to the idea of economies of scale—that grouping economic activity is more productive in many cases than spreading it out. For example, cities provide a large group of nearby customers, so that businesses can produce at an efficient economy of scale. They also provide a large group of workers and suppliers, so that business can hire easily and purchase whatever specialized inputs they need. Many of the attractions of cities,

like sports stadiums and museums, can operate only if they can draw on a large nearby population base. Cities are big enough to offer a wide variety of products, which is what many shoppers are looking for.

These factors are not exactly economies of scale in the narrow sense of the production function of a single firm, but they are related to growth in the overall size of population and market in an area. Cities are sometimes called “agglomeration economies.”

These agglomeration factors help to explain why every economy, as it develops, has an increasing proportion of its population living in urban areas. In the United States, about 80% of the population now lives in metropolitan areas (which include the suburbs around cities), compared to just 40% in 1900. However, in poorer nations of the world, including much of Africa, the proportion of the population in urban areas is only about 30%. One of the great challenges for these countries as their economies grow will be to manage the growth of the great cities that will arise.

If cities offer economic advantages that are a form of economies of scale, then why don't all or most people live in one giant city? At some point, agglomeration economies must turn into diseconomies. For example, traffic congestion may reach a point where the gains from being geographically nearby are counterbalanced by how long it takes to travel. High densities of people, cars, and factories can mean more garbage and air and water pollution. Facilities like parks or museums may become overcrowded. There may be economies of scale for negative activities like crime, because high densities of people and businesses, combined with the greater impersonality of cities, make it easier for illegal activities as well as legal ones. The future of cities, both in the United States and in other countries around the world, will be determined by their ability to benefit from the economies of agglomeration and to minimize or counterbalance the corresponding diseconomies.

A more common case is illustrated in [\[link\]](#) (b), where the LRAC curve has a flat-bottomed area of constant returns to scale. In this situation, any firm with a level of output between 5,000 and 20,000 will be able to produce at about the same level of average cost. Given that the market will demand one million dishwashers per year at a price of \$500, this market might have

as many as 200 producers (that is, one million dishwashers divided by firms making 5,000 each) or as few as 50 producers (one million dishwashers divided by firms making 20,000 each). The producers in this market will range in size from firms that make 5,000 units to firms that make 20,000 units. But firms that produce below 5,000 units or more than 20,000 will be unable to compete, because their average costs will be too high. Thus, if we see an industry where almost all plants are the same size, it is likely that the long-run average cost curve has a unique bottom point as in [\[link\]](#) (a). However, if the long-run average cost curve has a wide flat bottom like [\[link\]](#) (b), then firms of a variety of different sizes will be able to compete with each other.

The flat section of the long-run average cost curve in [\[link\]](#) (b) can be interpreted in two different ways. One interpretation is that a single manufacturing plant producing a quantity of 5,000 has the same average costs as a single manufacturing plant with four times as much capacity that produces a quantity of 20,000. The other interpretation is that one firm owns a single manufacturing plant that produces a quantity of 5,000, while another firm owns four separate manufacturing plants, which each produce a quantity of 5,000. This second explanation, based on the insight that a single firm may own a number of different manufacturing plants, is especially useful in explaining why the long-run average cost curve often has a large flat segment—and thus why a seemingly smaller firm may be able to compete quite well with a larger firm. At some point, however, the task of coordinating and managing many different plants raises the cost of production sharply, and the long-run average cost curve slopes up as a result.

In the examples to this point, the quantity demanded in the market is quite large (one million) compared with the quantity produced at the bottom of the long-run average cost curve (5,000, 10,000 or 20,000). In such a situation, the market is set for competition between many firms. But what if the bottom of the long-run average cost curve is at a quantity of 10,000 and the total market demand at that price is only slightly higher than that quantity—or even somewhat lower?

Return to [\[link\]](#) (a), where the bottom of the long-run average cost curve is at 10,000, but now imagine that the total quantity of dishwashers demanded in the market at that price of \$500 is only 30,000. In this situation, the total number of firms in the market would be three. A handful of firms in a market is called an “oligopoly,” and the chapter on [Monopolistic Competition and Oligopoly](#) will discuss the range of competitive strategies that can occur when oligopolies compete.

Alternatively, consider a situation, again in the setting of [\[link\]](#) (a), where the bottom of the long-run average cost curve is 10,000, but total demand for the product is only 5,000. (For simplicity, imagine that this demand is highly inelastic, so that it does not vary according to price.) In this situation, the market may well end up with a single firm—a monopoly—producing all 5,000 units. If any firm tried to challenge this monopoly while producing a quantity lower than 5,000 units, the prospective competitor firm would have a higher average cost, and so it would not be able to compete in the longer term without losing money. The chapter on [Monopoly](#) discusses the situation of a monopoly firm.

Thus, the shape of the long-run average cost curve reveals whether competitors in the market will be different sizes. If the LRAC curve has a single point at the bottom, then the firms in the market will be about the same size, but if the LRAC curve has a flat-bottomed segment of constant returns to scale, then firms in the market may be a variety of different sizes.

The relationship between the quantity at the minimum of the long-run average cost curve and the quantity demanded in the market at that price will predict how much competition is likely to exist in the market. If the quantity demanded in the market far exceeds the quantity at the minimum of the LRAC, then many firms will compete. If the quantity demanded in the market is only slightly higher than the quantity at the minimum of the LRAC, a few firms will compete. If the quantity demanded in the market is less than the quantity at the minimum of the LRAC, a single-producer monopoly is a likely outcome.

## **Shifting Patterns of Long-Run Average Cost**

New developments in production technology can shift the long-run average cost curve in ways that can alter the size distribution of firms in an industry.

For much of the twentieth century, the most common change has been to see alterations in technology, like the assembly line or the large department store, where large-scale producers seemed to gain an advantage over smaller ones. In the long-run average cost curve, the downward-sloping economies of scale portion of the curve stretched over a larger quantity of output.

However, new production technologies do not inevitably lead to a greater average size for firms. For example, in recent years some new technologies for generating electricity on a smaller scale have appeared. The traditional coal-burning electricity plants needed to produce 300 to 600 megawatts of power to exploit economies of scale fully. However, high-efficiency turbines to produce electricity from burning natural gas can produce electricity at a competitive price while producing a smaller quantity of 100 megawatts or less. These new technologies create the possibility for smaller companies or plants to generate electricity as efficiently as large ones. Another example of a technology-driven shift to smaller plants may be taking place in the tire industry. A traditional mid-size tire plant produces about six million tires per year. However, in 2000, the Italian company Pirelli introduced a new tire factory that uses many robots. The Pirelli tire plant produced only about one million tires per year, but did so at a lower average cost than a traditional mid-sized tire plant.

Controversy has simmered in recent years over whether the new information and communications technologies will lead to a larger or smaller size for firms. On one side, the new technology may make it easier for small firms to reach out beyond their local geographic area and find customers across a state, or the nation, or even across international boundaries. This factor might seem to predict a future with a larger number of small competitors. On the other side, perhaps the new information and communications technology will create “winner-take-all” markets where one large company will tend to command a large share of total sales, as Microsoft has done in the production of software for personal computers or Amazon has done in online bookselling. Moreover, improved information

and communication technologies might make it easier to manage many different plants and operations across the country or around the world, and thus encourage larger firms. This ongoing battle between the forces of smallness and largeness will be of great interest to economists, businesspeople, and policymakers.

**Note:**

**Amazon**

Traditionally, bookstores have operated in retail locations with inventories held either on the shelves or in the back of the store. These retail locations were very pricey in terms of rent. Amazon has no retail locations; it sells online and delivers by mail. Amazon offers almost any book in print, convenient purchasing, and prompt delivery by mail. Amazon holds its inventories in huge warehouses in low-rent locations around the world. The warehouses are highly computerized using robots and relatively low-skilled workers, making for low average costs per sale. Amazon demonstrates the significant advantages economies of scale can offer to a firm that exploits those economies.

## **Key Concepts and Summary**

A production technology refers to a specific combination of labor, physical capital, and technology that makes up a particular method of production.

In the long run, firms can choose their production technology, and so all costs become variable costs. In making this choice, firms will try to substitute relatively inexpensive inputs for relatively expensive inputs where possible, so as to produce at the lowest possible long-run average cost.

Economies of scale refers to a situation where as the level of output increases, the average cost decreases. Constant returns to scale refers to a situation where average cost does not change as output increases.

Diseconomies of scale refers to a situation where as output increases, average costs increase also.

The long-run average cost curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology. A downward-sloping LRAC shows economies of scale; a flat LRAC shows constant returns to scale; an upward-sloping LRAC shows diseconomies of scale. If the long-run average cost curve has only one quantity produced that results in the lowest possible average cost, then all of the firms competing in an industry should be the same size. However, if the LRAC has a flat segment at the bottom, so that a range of different quantities can be produced at the lowest average cost, the firms competing in the industry will display a range of sizes. The market demand in conjunction with the long-run average cost curve determines how many firms will exist in a given industry.

If the quantity demanded in the market of a certain product is much greater than the quantity found at the bottom of the long-run average cost curve, where the cost of production is lowest, the market will have many firms competing. If the quantity demanded in the market is less than the quantity at the bottom of the LRAC, there will likely be only one firm.

## Self-Check Questions

### Exercise:

#### Problem:

Return to the problem explained in [\[link\]](#) and [\[link\]](#). If the cost of labor remains at \$40, but the cost of a machine decreases to \$50, what would be the total cost of each method of production? Which method should the firm use, and why?

---

#### Solution:

The new table should look like this:

	<b>Labor Cost</b>	<b>Machine Cost</b>	<b>Total Cost</b>
Cost of technology 1	$10 \times \$40 = \$400$	$2 \times \$50 = \$100$	\$500
Cost of technology 2	$7 \times \$40 = \$280$	$4 \times \$50 = \$200$	\$480
Cost of technology 3	$3 \times \$40 = \$120$	$7 \times \$50 = \$350$	\$470

The firm should choose production technology 3 since it has the lowest total cost. This makes sense since, with cheaper machine hours, one would expect a shift in the direction of more machines and less labor.

### **Exercise:**

#### **Problem:**

Suppose the cost of machines increases to \$55, while the cost of labor stays at \$40. How would that affect the total cost of the three methods? Which method should the firm choose now?

---

#### **Solution:**

	<b>Labor Cost</b>	<b>Machine Cost</b>	<b>Total Cost</b>
Cost of technology 1	$10 \times \$40 = \$400$	$2 \times \$55 = \$110$	\$510

---

	<b>Labor Cost</b>	<b>Machine Cost</b>	<b>Total Cost</b>
Cost of technology 2	$7 \times \$40 = \$280$	$4 \times \$55 = \$220$	\$500
Cost of technology 3	$3 \times \$40 = \$120$	$7 \times \$55 = \$385$	\$505

The firm should choose production technology 2 since it has the lowest total cost. Because the cost of machines increased (relative to the previous question), you would expect a shift toward less capital and more labor.

### **Exercise:**

#### **Problem:**

Automobile manufacturing is an industry subject to significant economies of scale. Suppose there are four domestic auto manufacturers, but the demand for domestic autos is no more than 2.5 times the quantity produced at the bottom of the long-run average cost curve. What do you expect will happen to the domestic auto industry in the long run?

---

#### **Solution:**

This is the situation that existed in the United States in the 1970s. Since there is only demand enough for 2.5 firms to reach the bottom of the average cost curve, you would expect one firm will not be around in the long run, and at least one firm will be struggling.

## **Review Questions**

### **Exercise:**

**Problem:**

What shapes would you generally expect each of the following cost curves to have: fixed costs, variable costs, marginal costs, average total costs, and average variable costs?

**Exercise:**

**Problem:** What is a production technology?

**Exercise:**

**Problem:**

In choosing a production technology, how will firms react if one input becomes relatively more expensive?

**Exercise:**

**Problem:** What is a long-run average cost curve?

**Exercise:**

**Problem:**

What is the difference between economies of scale, constant returns to scale, and diseconomies of scale?

**Exercise:**

**Problem:**

What shape of a long-run average cost curve illustrates economies of scale, constant returns to scale, and diseconomies of scale?

**Exercise:**

**Problem:**

Why will firms in most markets be located at or close to the bottom of the long-run average cost curve?

## Critical Thinking Questions

### Exercise:

#### Problem:

It is clear that businesses operate in the short run, but do they ever operate in the long run? Discuss.

### Exercise:

#### Problem:

How would an improvement in technology, like the high-efficiency gas turbines or Pirelli tire plant, affect the long-run average cost curve of a firm? Can you draw the old curve and the new one on the same axes? How might such an improvement affect other firms in the industry?

### Exercise:

#### Problem:

Do you think that the taxicab industry in large cities would be subject to significant economies of scale? Why or why not?

## Problems

### Exercise:

#### Problem:

A small company that shovels sidewalks and driveways has 100 homes signed up for its services this winter. It can use various combinations of capital and labor: lots of labor with hand shovels, less labor with snow blowers, and still less labor with a pickup truck that has a snowplow on front. To summarize, the method choices are:

Method 1: 50 units of labor, 10 units of capital

Method 2: 20 units of labor, 40 units of capital

Method 3: 10 units of labor, 70 units of capital

If hiring labor for the winter costs \$100/unit and a unit of capital costs \$400, what production method should be chosen? What method should be chosen if the cost of labor rises to \$200/unit?

## **Glossary**

constant returns to scale

expanding all inputs proportionately does not change the average cost of production

diseconomies of scale

the long-run average cost of producing each individual unit increases as total output increases

long-run average cost (LRAC) curve

shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology

production technologies

alternative methods of combining inputs to produce output

short-run average cost (SRAC) curve

the average total cost curve in the short term; shows the total of the average fixed costs and the average variable costs

## Introduction to Perfect Competition

class="introduction"

Depending upon the competition and prices offered, a wheat farmer may choose to grow a different crop.

(Credit: modification of work by Daniel X. O'Neil/Flickr r Creative Commons)

**Note:****A Dime a Dozen**

When you were younger did you babysit, deliver papers, or mow the lawn for money? If so, you faced stiff competition from a lot of other competitors who offered identical services. There was nothing to stop others from offering their services too.

All of you charged the “going rate.” If you tried to charge more, your customers would simply buy from someone else. These conditions are very similar to the conditions agricultural growers face.

Growing a crop may be more difficult to start than a babysitting or lawn mowing service, but growers face the same fierce competition. In the grand scale of world agriculture, farmers face competition from thousands of others because they sell an identical product. After all, winter wheat is winter wheat. But it is relatively easy for farmers to leave the marketplace for another crop. In this case, they do not sell the family farm, they switch crops.

Take the case of the upper Midwest region of the United States—for many generations the area was called “King Wheat.” According to the United

States Department of Agriculture National Agricultural Statistics Service, statistics by state, in 1997, 11.6 million acres of wheat and 780,000 acres of corn were planted in North Dakota. In the intervening 15 or so years has the mix of crops changed? Since it is relatively easy to switch crops, did farmers change what was planted as the relative crop prices changed? We will find out at chapter's end.

In the meantime, let's consider the topic of this chapter—the perfectly competitive market. This is a market in which entry and exit are relatively easy and competitors are “a dime a dozen.”

**Note:**

**Introduction to Perfect Competition**

In this chapter, you will learn about:

- Perfect Competition and Why It Matters
- How Perfectly Competitive Firms Make Output Decisions
- Entry and Exit Decisions in the Long Run
- Efficiency in Perfectly Competitive Markets

All businesses face two realities: no one is required to buy their products, and even customers who might want those products may buy from other businesses instead. Firms that operate in perfectly competitive markets face this reality. In this chapter, you will learn how such firms make decisions about how much to produce, how much profit they make, whether to stay in business or not, and many others. Industries differ from one another in terms of how many sellers there are in a specific market, how easy or difficult it is for a new firm to enter, and the type of products that are sold. This is referred to as the **market structure** of the industry. In this chapter, we focus on perfect competition. However, in other chapters we will examine other industry types: [Monopoly](#) and [Monopolistic Competition and Oligopoly](#).

## Perfect Competition and Why It Matters

By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are said to be in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A perfectly competitive firm is known as a **price taker**, because the pressure of competing firms forces them to accept the prevailing equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a penny, it will lose all of its sales to competitors. When a wheat grower, as discussed in the Bring it Home feature, wants to know what the going price of wheat is, he or she has to go to the computer or listen to the radio to check. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is a hypothetical extreme; however, producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. According to the United States Department of Agriculture monthly reports, in 2015, U.S. corn farmers received an average price of \$6.00 per bushel and wheat farmers received an average price of \$6.00 per bushel. A corn farmer who attempted to sell at \$7.00 per bushel, or a wheat grower who attempted to sell for \$8.00 per bushel, would not have found any buyers. A perfectly competitive

firm will not sell below the equilibrium price either. Why should they when they can sell all they want at the higher price? Other examples of agricultural markets that operate in close to perfectly competitive markets are small roadside produce markets and small organic farmers.

**Note:**

Visit this [website](#) that reveals the current value of various commodities.



This chapter examines how profit-seeking firms decide how much to produce in perfectly competitive markets. Such firms will analyze their costs as discussed in the chapter on [Cost and Industry Structure](#). In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest. In this example, the “short run” refers to a situation in which firms are producing with one fixed input and incur fixed costs of production. (In the real world, firms can have many fixed inputs.)

In the long run, perfectly competitive firms will react to profits by increasing production. They will respond to losses by reducing production or exiting the market. Ultimately, a long-run *equilibrium* will be attained when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

## Key Concepts and Summary

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products. A short-run production period is when firms are producing with some fixed inputs. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry and seller profits are driven to zero.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

## **Self-Check Questions**

### **Exercise:**

#### **Problem:**

Firms in a perfectly competitive market are said to be “price takers”—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?

---

#### **Solution:**

No, you would not raise the price. Your product is exactly the same as the product of the many other firms in the market. If your price is greater than that of your competitors, then your customers would switch to them and stop buying from you. You would lose all your sales.

### **Exercise:**

**Problem:**

Would independent trucking fit the characteristics of a perfectly competitive industry?

---

**Solution:**

Possibly. Independent truckers are by definition small and numerous. All that is required to get into the business is a truck (not an inexpensive asset, though) and a commercial driver's license. To exit, one need only sell the truck. All trucks are essentially the same, providing transportation from point A to point B. (We're assuming we not talking about specialized trucks.) Independent truckers must take the going rate for their service, so independent trucking does seem to have most of the characteristics of perfect competition.

**Review Questions****Exercise:****Problem:**

A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How "small" is "small"?

**Exercise:****Problem:**

What are the four basic assumptions of perfect competition? Explain in words what they imply for a perfectly competitive firm.

**Exercise:**

**Problem:** What is a "price taker" firm?

## Critical Thinking Questions

### Exercise:

#### Problem:

Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?

### Exercise:

#### Problem:

Can you name five examples of perfectly competitive markets? Why or why not?

## Glossary

market structure

the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition

each firm faces many competitors that sell identical products

price taker

a firm in a perfectly competitive market that must take the prevailing market price as given

## How Perfectly Competitive Firms Make Output Decisions

By the end of this section, you will be able to:

- Calculate profits by comparing total revenue and total cost
- Identify profits and losses with the average cost curve
- Explain the shutdown point
- Determine the price at which a firm should continue producing in the short run

A perfectly competitive firm has only one major decision to make—namely, what quantity to produce. To understand why this is so, consider a different way of writing out the basic definition of profit:

### Equation:

$$\begin{aligned}\text{Profit} &= \text{Total revenue} - \text{Total cost} \\ &= (\text{Price})(\text{Quantity produced}) - (\text{Average cost})(\text{Quantity produced})\end{aligned}$$

Since a perfectly competitive firm must accept the price for its output as determined by the product's market demand and supply, it cannot choose the price it charges. This is already determined in the profit equation, and so the perfectly competitive firm can sell any number of units at exactly the same price. It implies that the firm faces a perfectly elastic demand curve for its product: buyers are willing to buy any number of units of output from the firm at the market price. When the perfectly competitive firm chooses what quantity to produce, then this quantity—along with the prices prevailing in the market for output and inputs—will determine the firm's total revenue, total costs, and ultimately, level of profits.

## Determining the Highest Profit by Comparing Total Revenue and Total Cost

A perfectly competitive firm can sell as large a quantity as it wishes, as long as it accepts the prevailing market price. Total revenue is going to increase as the firm sells more, depending on the price of the product and the number of units sold. If you increase the number of units sold at a given price, then total revenue will increase. If the price of the product increases for every unit sold, then total revenue also increases. As an example of how a perfectly competitive firm decides what quantity to produce, consider the case of a small farmer who produces raspberries and sells them frozen for \$4 per pack. Sales of one pack of raspberries will bring in \$4, two packs will be \$8, three packs will be \$12, and so on. If, for example, the price of frozen raspberries doubles to \$8 per pack, then sales of one pack of raspberries will be \$8, two packs will be \$16, three packs will be \$24, and so on.

Total revenue and total costs for the raspberry farm, broken down into fixed and variable costs, are shown in [\[link\]](#) and also appear in [\[link\]](#). The horizontal axis shows the quantity of frozen raspberries produced in packs; the vertical axis shows both total revenue and total costs, measured in dollars. The total cost curve intersects with the vertical axis at a value that shows the level of fixed costs, and then slopes upward. All these cost curves follow the same characteristics as the curves covered in the [Cost and Industry Structure](#) chapter.

Total Cost and Total Revenue at the Raspberry Farm



Total revenue for a perfectly competitive firm is a straight line

sloping up. The slope is equal to the price of the good. Total cost also slopes up, but with some curvature. At higher levels of output, total cost begins to slope upward more steeply because of diminishing marginal returns. The maximum profit will occur at the quantity where the gap of total revenue over total cost is largest.

Quantity (Q)	Total Cost (TC)	Fixed Cost (FC)	Variable Cost (VC)	Total Revenue (TR)	Profit
0	\$62	\$62	-	\$0	-\$62
10	\$90	\$62	\$28	\$40	-\$50
20	\$110	\$62	\$48	\$80	-\$30
30	\$126	\$62	\$64	\$120	-\$6
40	\$144	\$62	\$82	\$160	\$16
50	\$166	\$62	\$104	\$200	\$34
60	\$192	\$62	\$130	\$240	\$48
70	\$224	\$62	\$162	\$280	\$56
80	\$264	\$62	\$202	\$320	\$56
90	\$324	\$62	\$262	\$360	\$36
100	\$404	\$62	\$342	\$400	-\$4

#### Total Cost and Total Revenue at the Raspberry Farm

Based on its total revenue and total cost curves, a perfectly competitive firm like the raspberry farm can calculate the quantity of output that will provide the highest level of profit. At any given quantity, total revenue minus total cost will equal profit. One way to determine the most profitable quantity to produce is to see at what quantity total revenue exceeds total cost by the largest amount. On [\[link\]](#), the vertical gap between total revenue and total cost represents either profit (if total revenues are greater than total costs at a certain quantity) or losses (if total costs are greater than total revenues at a certain quantity). In this example, total costs will exceed total revenues at output levels from 0 to 40, and so over this range of output, the firm will be making losses. At output levels from 50 to 80, total revenues exceed total costs, so the firm is earning profits. But then at an output of 90 or 100, total costs again exceed total revenues and the firm is making losses. Total profits appear in the final column of [\[link\]](#). The highest total profits in the table, as in the figure that is based on the table values, occur at an output of 70–80, when profits will be \$56.

A higher price would mean that total revenue would be higher for every quantity sold. A lower price would mean that total revenue would be lower for every quantity sold. What happens if the price drops low enough so that the total revenue line is completely below the total cost curve; that is, at every level of output, total costs are higher than total revenues? In this instance, the best the firm can do is to suffer losses. But a profit-maximizing firm will

prefer the quantity of output where total revenues come closest to total costs and thus where the losses are smallest.

(Later we will see that sometimes it will make sense for the firm to shutdown, rather than stay in operation producing output.)

### Comparing Marginal Revenue and Marginal Costs

Firms often do not have the necessary data they need to draw a complete total cost curve for all levels of production. They cannot be sure of what total costs would look like if they, say, doubled production or cut production in half, because they have not tried it. Instead, firms experiment. They produce a slightly greater or lower quantity and observe how profits are affected. In economic terms, this practical approach to maximizing profits means looking at how changes in production affect marginal revenue and marginal cost.

[\[link\]](#) presents the marginal revenue and marginal cost curves based on the total revenue and total cost in [\[link\]](#). The **marginal revenue** curve shows the additional revenue gained from selling one more unit. As mentioned before, a firm in perfect competition faces a perfectly elastic demand curve for its product—that is, the firm’s demand curve is a horizontal line drawn at the market price level. This also means that the firm’s marginal revenue curve is the same as the firm’s demand curve: Every time a consumer demands one more unit, the firm sells one more unit and revenue goes up by exactly the same amount equal to the market price. In this example, every time a pack of frozen raspberries is sold, the firm’s revenue increases by \$4. [\[link\]](#) shows an example of this. This condition only holds for price taking firms in perfect competition where:

**Equation:**

$$\text{marginal revenue} = \text{price}$$

The formula for marginal revenue is:

**Equation:**

$$\text{marginal revenue} = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

Price	Quantity	Total Revenue	Marginal Revenue
\$4	1	\$4	-
\$4	2	\$8	\$4
\$4	3	\$12	\$4
\$4	4	\$16	\$4

Notice that marginal revenue does not change as the firm produces more output. That is because the price is determined by supply and demand and does not change as the farmer produces more (keeping in mind that, due to the relative small size of each firm, increasing their supply has no impact on the total market supply where price is determined).

Since a perfectly competitive firm is a price taker, it can sell whatever quantity it wishes at the market-determined price. Marginal cost, the cost per additional unit sold, is calculated by dividing the change in total cost by the

change in quantity. The formula for marginal cost is:

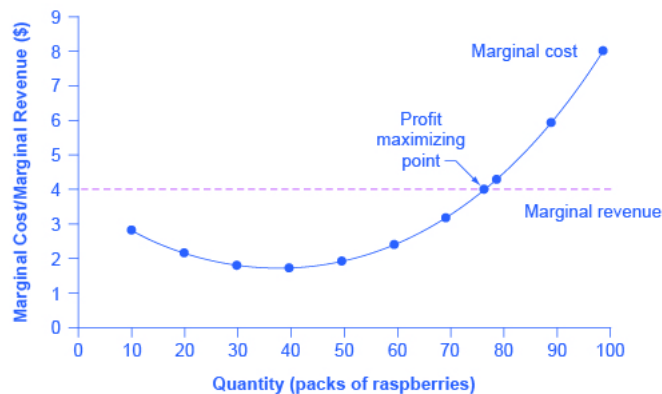
**Equation:**

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}}$$

Ordinarily, marginal cost changes as the firm produces a greater quantity.

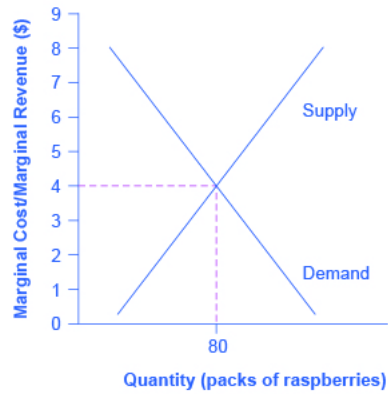
In the raspberry farm example, shown in [\[link\]](#), [\[link\]](#) and [\[link\]](#), marginal cost at first declines as production increases from 10 to 20 to 30 packs of raspberries—which represents the area of increasing marginal returns that is not uncommon at low levels of production. But then marginal costs start to increase, displaying the typical pattern of diminishing marginal returns. If the firm is producing at a quantity where  $MR > MC$ , like 40 or 50 packs of raspberries, then it can increase profit by increasing output because the marginal revenue is exceeding the marginal cost. If the firm is producing at a quantity where  $MC > MR$ , like 90 or 100 packs, then it can increase profit by reducing output because the reductions in marginal cost will exceed the reductions in marginal revenue. The firm's profit-maximizing choice of output will occur where  $MR = MC$  (or at a choice close to that point). You will notice that what occurs on the production side is exemplified on the cost side. This is referred to as duality.

**Marginal Revenues and Marginal Costs at the Raspberry Farm: Individual Farmer**



For a perfectly competitive firm, the marginal revenue (MR) curve is a horizontal straight line because it is equal to the price of the good, which is determined by the market, shown in [\[link\]](#). The marginal cost (MC) curve is sometimes first downward-sloping, if there is a region of increasing marginal returns at low levels of output, but is eventually upward-sloping at higher levels of output as diminishing marginal returns kick in.

**Marginal Revenues and Marginal Costs at the Raspberry Farm: Raspberry Market**



The equilibrium price of raspberries is determined through the interaction of market supply and market demand at \$4.00.

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	\$62	\$62	-	-	-	-
10	\$90	\$62	\$28	\$2.80	\$40	\$4.00
20	\$110	\$62	\$48	\$2.00	\$80	\$4.00
30	\$126	\$62	\$64	\$1.60	\$120	\$4.00
40	\$144	\$62	\$82	\$1.80	\$160	\$4.00
50	\$166	\$62	\$104	\$2.20	\$200	\$4.00
60	\$192	\$62	\$130	\$2.60	\$240	\$4.00
70	\$224	\$62	\$162	\$3.20	\$280	\$4.00
80	\$264	\$62	\$202	\$4.00	\$320	\$4.00
90	\$324	\$62	\$262	\$6.00	\$360	\$4.00
100	\$404	\$62	\$342	\$8.00	\$400	\$4.00

#### Marginal Revenues and Marginal Costs at the Raspberry Farm

In this example, the marginal revenue and marginal cost curves cross at a price of \$4 and a quantity of 80 produced. If the farmer started out producing at a level of 60, and then experimented with increasing production to 70, marginal revenues from the increase in production would exceed marginal costs—and so profits would rise. The farmer has an incentive to keep producing. From a level of 70 to 80, marginal cost and marginal revenue are

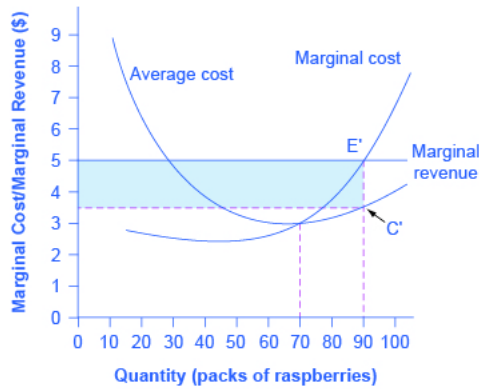
equal so profit doesn't change. If the farmer then experimented further with increasing production from 80 to 90, he would find that marginal costs from the increase in production are greater than marginal revenues, and so profits would decline.

The profit-maximizing choice for a perfectly competitive firm will occur where marginal revenue is equal to marginal cost—that is, where  $MR = MC$ . A profit-seeking firm should keep expanding production as long as  $MR > MC$ . But at the level of output where  $MR = MC$ , the firm should recognize that it has achieved the highest possible level of economic profits. (In the example above, the profit maximizing output level is between 70 and 80 units of output, but the firm will not know they've maximized profit until they reach 80, where  $MR = MC$ .) Expanding production into the zone where  $MR < MC$  will only reduce economic profits. Because the marginal revenue received by a perfectly competitive firm is equal to the price  $P$ , so that  $P = MR$ , the profit-maximizing rule for a perfectly competitive firm can also be written as a recommendation to produce at the quantity where  $P = MC$ .

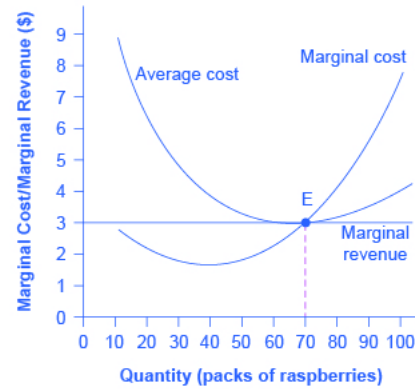
### **Profits and Losses with the Average Cost Curve**

Does maximizing profit (producing where  $MR = MC$ ) imply an actual economic profit? The answer depends on the relationship between price and average total cost. If the price that a firm charges is higher than its average cost of production for that quantity produced, then the firm will earn profits. Conversely, if the price that a firm charges is lower than its average cost of production, the firm will suffer losses. You might think that, in this situation, the farmer may want to shut down immediately. Remember, however, that the firm has already paid for fixed costs, such as equipment, so it may continue to produce and incur a loss. [\[link\]](#) illustrates three situations: (a) where price intersects marginal cost at a level above the average cost curve, (b) where price intersects marginal cost at a level equal to the average cost curve, and (c) where price intersects marginal cost at a level below the average cost curve.

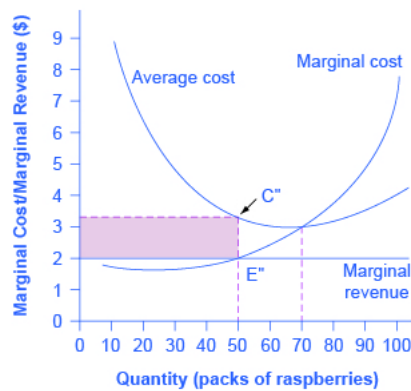
Price and Average Cost at the Raspberry Farm



(a) Price is above average cost



(b) Price equals cost



(c) Price is below average cost

In (a), price intersects marginal cost above the average cost curve. Since price is greater than average cost, the firm is making a profit. In (b), price intersects marginal cost at the minimum point of the average cost curve. Since price is equal to average cost, the firm is breaking even. In (c), price intersects marginal cost below the average cost curve. Since price is less than average cost, the firm is making a loss.

First consider a situation where the price is equal to \$5 for a pack of frozen raspberries. The rule for a profit-maximizing perfectly competitive firm is to produce the level of output where  $\text{Price} = \text{MR} = \text{MC}$ , so the raspberry farmer will produce a quantity of 90, which is labeled as [e](#) in [\[link\]](#) (a). Remember that the area of a rectangle is equal to its base multiplied by its height. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 90 packs (the base) up to point E' (the height), over to the price of \$5, and back to the origin. The average cost of producing 80 packs is shown by point C or about \$3.50. Total costs will be the quantity of 80 times the average cost of \$3.50, which is shown by the area of the rectangle from the origin to a quantity of 90, up to point C, over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is greater than total cost. Thus, profits will be the blue shaded rectangle on top.

It can be calculated as:

**Equation:**

$$\begin{aligned}
 \text{profit} &= \text{total revenue} - \text{total cost} \\
 &= (90)(\$5.00) - (90)(\$3.50) \\
 &= \$135
 \end{aligned}$$

Or, it can be calculated as:

**Equation:**

$$\begin{aligned}
 \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\
 &= (\$5.00 - \$3.50) \times 90 \\
 &= \$135
 \end{aligned}$$

Now consider [\[link\]](#) (b), where the price has fallen to \$3.00 for a pack of frozen raspberries. Again, the perfectly competitive firm will choose the level of output where Price = MR = MC, but in this case, the quantity produced will be 70. At this price and output level, where the marginal cost curve is crossing the average cost curve, the price received by the firm is exactly equal to its average cost of production.

The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 70 packs (the base) up to point E (the height), over to the price of \$3, and back to the origin. The average cost of producing 70 packs is shown by point C'. Total costs will be the quantity of 70 times the average cost of \$3.00, which is shown by the area of the rectangle from the origin to a quantity of 70, up to point E, over to the vertical axis and down to the origin. It should be clear from that the rectangles for total revenue and total cost are the same. Thus, the firm is making zero profit. The calculations are as follows:

**Equation:**

$$\begin{aligned}
 \text{profit} &= \text{total revenue} - \text{total cost} \\
 &= (70)(\$3.00) - (70)(\$3.00) \\
 &= \$0
 \end{aligned}$$

Or, it can be calculated as:

**Equation:**

$$\begin{aligned}
 \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\
 &= (\$3.00 - \$3.00) \times 70 \\
 &= \$0
 \end{aligned}$$

In [\[link\]](#) (c), the market price has fallen still further to \$2.00 for a pack of frozen raspberries. At this price, marginal revenue intersects marginal cost at a quantity of 50. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 50 packs (the base) up to point E" (the height), over to the price of \$2, and back to the origin. The average cost of producing 50 packs is shown by point C" or about \$3.30. Total costs will be the quantity of 50 times the average cost of \$3.30, which is shown by the area of the rectangle from the origin to a quantity of 50, up to point C", over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is less than total cost. Thus, the firm is losing money and the loss (or negative profit) will be the rose-shaded rectangle.

The calculations are:

**Equation:**

$$\begin{aligned}
 \text{profit} &= (\text{total revenue} - \text{total cost}) \\
 &= (50)(\$2.00) - (50)(\$3.30) \\
 &= -\$77.50
 \end{aligned}$$

Or:

**Equation:**

$$\begin{aligned}\text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\$1.75 - \$3.30) \times 50 \\ &= -\$77.50\end{aligned}$$

If the market price received by a perfectly competitive firm leads it to produce at a quantity where the price is greater than average cost, the firm will earn profits. If the price received by the firm causes it to produce at a quantity where price equals average cost, which occurs at the minimum point of the AC curve, then the firm earns zero profits. Finally, if the price received by the firm leads it to produce at a quantity where the price is less than average cost, the firm will earn losses. This is summarized in [\[link\]](#).

If...	Then...
Price > ATC	Firm earns an economic profit
Price = ATC	Firm earns zero economic profit
Price < ATC	Firm earns a loss

## The Shutdown Point

The possibility that a firm may earn losses raises a question: Why can the firm not avoid losses by shutting down and not producing at all? The answer is that shutting down can reduce variable costs to zero, but in the short run, the firm has already paid for fixed costs. As a result, if the firm produces a quantity of zero, it would still make losses because it would still need to pay for its fixed costs. So, when a firm is experiencing losses, it must face a question: should it continue producing or should it shut down?

As an example, consider the situation of the Yoga Center, which has signed a contract to rent space that costs \$10,000 per month. If the firm decides to operate, its marginal costs for hiring yoga teachers is \$15,000 for the month. If the firm shuts down, it must still pay the rent, but it would not need to hire labor. [\[link\]](#) shows three possible scenarios. In the first scenario, the Yoga Center does not have any clients, and therefore does not make any revenues, in which case it faces losses of \$10,000 equal to the fixed costs. In the second scenario, the Yoga Center has clients that earn the center revenues of \$10,000 for the month, but ultimately experiences losses of \$15,000 due to having to hire yoga instructors to cover the classes. In the third scenario, the Yoga Center earns revenues of \$20,000 for the month, but experiences losses of \$5,000.

In all three cases, the Yoga Center loses money. In all three cases, when the rental contract expires in the long run, assuming revenues do not improve, the firm should exit this business. In the short run, though, the decision varies depending on the level of losses and whether the firm can cover its variable costs. In scenario 1, the center does not have any revenues, so hiring yoga teachers would increase variable costs and losses, so it should shut down and only incur its fixed costs. In scenario 2, the center's losses are greater because it does not make enough revenue to offset the increased variable costs plus fixed costs, so it should shut down immediately. If price is below the minimum average variable cost, the firm must shut down. In contrast, in scenario 3 the revenue that the center can earn is high enough that the losses diminish when it remains open, so the center should remain open in the short run.

**Scenario 1**

If the center shuts down now, revenues are zero but it will not incur any variable costs and would only need to pay fixed costs of \$10,000.

**Equation:**

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= 0 - \$10,000 \\ &= -\$10,000\end{aligned}$$

**Scenario 2**

The center earns revenues of \$10,000, and variable costs are \$15,000. The center should shut down now.

**Equation:**

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= \$10,000 - (\$10,000 + \$15,000) \\ &= -\$15,000\end{aligned}$$

**Scenario 3**

The center earns revenues of \$20,000, and variable costs are \$15,000. The center should continue in business.

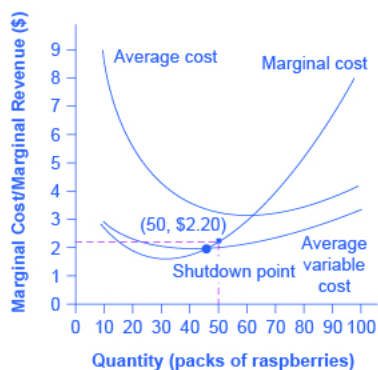
**Equation:**

$$\begin{aligned}\text{profit} &= \text{total revenue} - (\text{fixed costs} + \text{variable cost}) \\ &= \$20,000 - (\$10,000 + \$15,000) \\ &= -\$5,000\end{aligned}$$

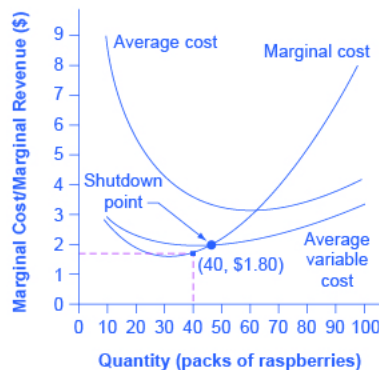
**Should the Yoga Center Shut Down Now or Later?**

This example suggests that the key factor is whether a firm can earn enough revenues to cover at least its variable costs by remaining open. Let's return now to our raspberry farm. [\[link\]](#) illustrates this lesson by adding the average variable cost curve to the marginal cost and average cost curves. At a price of \$2.20 per pack, as shown in [\[link\]](#) (a), the farm produces at a level of 50. It is making losses of \$56 (as explained earlier), but price is above average variable cost and so the firm continues to operate. However, if the price declined to \$1.80 per pack, as shown in [\[link\]](#) (b), and if the firm applied its rule of producing where  $P = MR = MC$ , it would produce a quantity of 40. This price is below average variable cost for this level of output. If the farmer cannot pay workers (the variable costs), then it has to shut down. At this price and output, total revenues would be \$72 (quantity of 40 times price of \$1.80) and total cost would be \$144, for overall losses of \$72. If the farm shuts down, it must pay only its fixed costs of \$62, so shutting down is preferable to selling at a price of \$1.80 per pack.

**The Shutdown Point for the Raspberry Farm**



(a) Price is above average variable cost



(b) Price is below average variable cost

In (a), the farm produces at a level of 50. It is making losses of \$56, but price is above average variable cost, so it continues to operate. In (b), total revenues are \$72 and total cost is \$144, for overall losses of \$72. If the farm shuts down, it must pay only its fixed costs of \$62. Shutting down is preferable to selling at a price of \$1.80 per pack.

Looking at [\[link\]](#), if the price falls below \$2.05, the minimum average variable cost, the firm must shut down.

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Average Cost	Average Variable Cost
0	\$62	\$62	-	-	-	-
10	\$90	\$62	\$28	\$2.80	\$9.00	\$2.80
20	\$110	\$62	\$48	\$2.00	\$5.50	\$2.40
30	\$126	\$62	\$64	\$1.60	\$4.20	\$2.13
40	\$144	\$62	\$82	\$1.80	\$3.60	\$2.05
50	\$166	\$62	\$104	\$2.20	\$3.32	\$2.08
60	\$192	\$62	\$130	\$2.60	\$3.20	\$2.16
70	\$224	\$62	\$162	\$3.20	\$3.20	\$2.31
80	\$264	\$62	\$202	\$4.00	\$3.30	\$2.52
90	\$324	\$62	\$262	\$6.00	\$3.60	\$2.91
100	\$404	\$62	\$342	\$8.00	\$4.04	\$3.42

Cost of Production for the Raspberry Farm

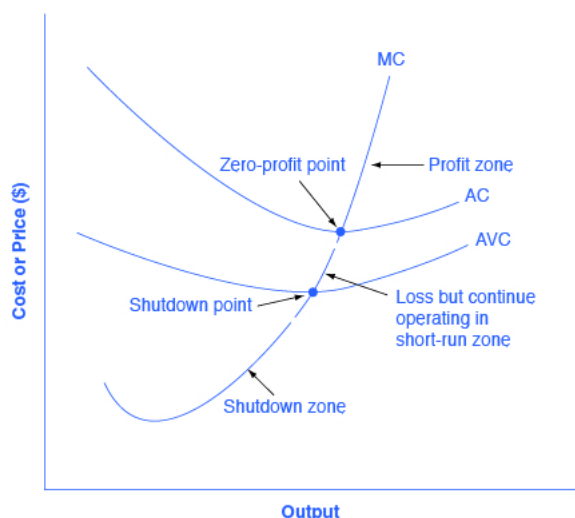
The intersection of the average variable cost curve and the marginal cost curve, which shows the price where the firm would lack enough revenue to cover its variable costs, is called the **shutdown point**. If the perfectly competitive firm can charge a price above the shutdown point, then the firm is at least covering its average variable costs. It is also making enough revenue to cover at least a portion of fixed costs, so it should limp ahead even if it is making losses in the short run, since at least those losses will be smaller than if the firm shuts down immediately and incurs a loss equal to total fixed costs. However, if the firm is receiving a price below the price at the shutdown point, then the firm is not even covering its variable costs. In this case, staying open is making the firm's losses larger, and it should shut down immediately. To summarize, if:

- price < minimum average variable cost, then firm shuts down
- price = minimum average variable cost, then firm stays in business

## Short-Run Outcomes for Perfectly Competitive Firms

The average cost and average variable cost curves divide the marginal cost curve into three segments, as shown in [\[link\]](#). At the market price, which the perfectly competitive firm accepts as given, the profit-maximizing firm chooses the output level where price or marginal revenue, which are the same thing for a perfectly competitive firm, is equal to marginal cost:  $P = MR = MC$ .

Profit, Loss, Shutdown



The marginal cost curve can be divided into three zones, based on where it is crossed by the average cost and average variable cost curves. The point where MC crosses AC is called the zero-profit point. If the firm is operating at a level of output where the market price is at a level higher than the zero-profit point, then price will be greater than average cost and the firm is earning profits. If the price is exactly at the zero-profit point, then the firm is making zero profits. If price falls in the zone between the shutdown point and the zero-profit point, then the firm is making losses but will continue to operate in the short run, since it is covering its variable costs. However, if price falls below the price at the shutdown point, then the firm will shut down immediately, since it is not even covering its variable costs.

First consider the upper zone, where prices are above the level where marginal cost (MC) crosses average cost (AC) at the zero profit point. At any price above that level, the firm will earn profits in the short run. If the price

falls exactly on the zero profit point where the MC and AC curves cross, then the firm earns zero profits. If a price falls into the zone between the zero profit point, where MC crosses AC, and the shutdown point, where MC crosses AVC, the firm will be making losses in the short run—but since the firm is more than covering its variable costs, the losses are smaller than if the firm shut down immediately. Finally, consider a price at or below the shutdown point where MC crosses AVC. At any price like this one, the firm will shut down immediately, because it cannot even cover its variable costs.

## Marginal Cost and the Firm's Supply Curve

For a perfectly competitive firm, the marginal cost curve is identical to the firm's supply curve starting from the minimum point on the average variable cost curve. To understand why this perhaps surprising insight holds true, first think about what the supply curve means. A firm checks the market price and then looks at its supply curve to decide what quantity to produce. Now, think about what it means to say that a firm will maximize its profits by producing at the quantity where  $P = MC$ . This rule means that the firm checks the market price, and then looks at its marginal cost to determine the quantity to produce—and makes sure that the price is greater than the minimum average variable cost. In other words, the marginal cost curve above the minimum point on the average variable cost curve becomes the firm's supply curve.

### Note:

Watch this [video](#) that addresses how drought in the United States can impact food prices across the world. (Note that the story on the drought is the second one in the news report; you need to let the video play through the first story in order to watch the story on the drought.)



As discussed in the chapter on [Demand and Supply](#), many of the reasons that supply curves shift relate to underlying changes in costs. For example, a lower price of key inputs or new technologies that reduce production costs cause supply to shift to the right; in contrast, bad weather or added government regulations can add to costs of certain goods in a way that causes supply to shift to the left. These shifts in the firm's supply curve can also be interpreted as shifts of the marginal cost curve. A shift in costs of production that increases marginal costs at all levels of output—and shifts MC to the left—will cause a perfectly competitive firm to produce less at any given market price. Conversely, a shift in costs of production that decreases marginal costs at all levels of output will shift MC to the right and as a result, a competitive firm will choose to expand its level of output at any given price. The following Work It Out feature will walk you through an example.

### Note:

#### At What Price Should the Firm Continue Producing in the Short Run?

To determine the short-run economic condition of a firm in perfect competition, follow the steps outlined below. Use the data shown in [\[link\]](#).

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	\$28	\$20	\$0	-	-	-	-	-	-
1	\$28	\$20	\$20	-	-	-	-	-	-
2	\$28	\$20	\$25	-	-	-	-	-	-
3	\$28	\$20	\$35	-	-	-	-	-	-
4	\$28	\$20	\$52	-	-	-	-	-	-
5	\$28	\$20	\$80	-	-	-	-	-	-

Step 1. Determine the cost structure for the firm. For a given total fixed costs and variable costs, calculate total cost, average variable cost, average total cost, and marginal cost. Follow the formulas given in the [Cost and Industry Structure](#) chapter. These calculations are shown in [\[link\]](#).

Q	P	TFC	TVC	TC (TFC+TVC)	AVC (TVC/Q)	ATC (TC/Q)	MC ( $TC_2 - TC_1$ ) / ( $Q_2 - Q_1$ )
0	\$28	\$20	\$0	$\$20 + \$0 = \$20$	-	-	-
1	\$28	\$20	\$20	$\$20 + \$20 = \$40$	$\$20 / 1 = \$20.00$	$\$40 / 1 = \$40.00$	$(\$40 - \$20) / (1 - 0) = \$20$
2	\$28	\$20	\$25	$\$20 + \$25 = \$45$	$\$25 / 2 = \$12.50$	$\$45 / 2 = \$22.50$	$(\$45 - \$40) / (2 - 1) = \$5$
3	\$28	\$20	\$35	$\$20 + \$35 = \$55$	$\$35 / 3 = \$11.67$	$\$55 / 3 = \$18.33$	$(\$55 - \$45) / (3 - 2) = \$10$
4	\$28	\$20	\$52	$\$20 + \$52 = \$72$	$\$52 / 4 = \$13.00$	$\$72 / 4 = \$18.00$	$(\$72 - \$55) / (4 - 3) = \$17$
5	\$28	\$20	\$80	$\$20 + \$80 = \$100$	$\$80 / 5 = \$16.00$	$\$100 / 5 = \$20.00$	$(\$100 - \$72) / (5 - 4) = \$28$

Step 2. Determine the market price that the firm receives for its product. This should be given information, as the firm in perfect competition is a price taker. With the given price, calculate total revenue as equal to price multiplied by quantity for all output levels produced. In this example, the given price is \$28. You can see that in the second column of [\[link\]](#).

Quantity	Price	Total Revenue (P × Q)
0	\$28	\$28×0=\$0
1	\$28	\$28×1=\$28
2	\$28	\$28×2=\$56
3	\$28	\$28×3=\$84
4	\$28	\$28×4=\$112
5	\$28	\$28×5=\$140

Step 3. Calculate profits as total cost subtracted from total revenue, as shown in [\[link\]](#).

Quantity	Total Revenue	Total Cost	Profits (TR–TC)
0	\$0	\$20	\$0–\$20=–\$20
1	\$28	\$40	\$28–\$40=–\$12
2	\$56	\$45	\$56–\$45=\$11
3	\$84	\$55	\$84–\$55=\$29
4	\$112	\$72	\$112–\$72=\$40
5	\$140	\$100	\$140–\$100=\$40

Step 4. To find the profit-maximizing output level, look at the Marginal Cost column (at every output level produced), as shown in [\[link\]](#), and determine where it is equal to the market price. The output level where price equals the marginal cost is the output level that maximizes profits.

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	\$28	\$20	\$0	\$20	-	-	-	\$0	–\$20
1	\$28	\$20	\$20	\$40	\$20.00	\$40.00	\$20	\$28	–\$12
2	\$28	\$20	\$25	\$45	\$12.50	\$22.50	\$5	\$56	\$11
3	\$28	\$20	\$35	\$55	\$11.67	\$18.33	\$10	\$84	\$29

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
4	\$28	\$20	\$52	\$72	\$13.00	\$18.00	\$17	\$112	\$40
5	<b>\$28</b>	\$20	\$80	\$100	\$16.40	\$20.40	<b>\$30</b>	\$140	<b>\$40</b>

Step 5. Once you have determined the profit-maximizing output level (in this case, output quantity 5), you can look at the amount of profits made (in this case, \$40).

Step 6. If the firm is making economic losses, the firm needs to determine whether it produces the output level where price equals marginal revenue and equals marginal cost or it shuts down and only incurs its fixed costs.

Step 7. For the output level where marginal revenue is equal to marginal cost, check if the market price is greater than the average variable cost of producing that output level.

- If  $P > AVC$  but  $P < ATC$ , then the firm continues to produce in the short-run, making economic losses.
- If  $P < AVC$ , then the firm stops producing and only incurs its fixed costs.

In this example, the price of \$28 is greater than the AVC (\$16.40) of producing 5 units of output, so the firm continues producing.

## Key Concepts and Summary

As a perfectly competitive firm produces a greater quantity of output, its total revenue steadily increases at a constant rate determined by the given market price. Profits will be highest (or losses will be smallest) at the quantity of output where total revenues exceed total costs by the greatest amount (or where total revenues fall short of total costs by the smallest amount). Alternatively, profits will be highest where marginal revenue, which is price for a perfectly competitive firm, is equal to marginal cost. If the market price faced by a perfectly competitive firm is above average cost at the profit-maximizing quantity of output, then the firm is making profits. If the market price is below average cost at the profit-maximizing quantity of output, then the firm is making losses.

If the market price is equal to average cost at the profit-maximizing level of output, then the firm is making zero profits. The point where the marginal cost curve crosses the average cost curve, at the minimum of the average cost curve, is called the “zero profit point.” If the market price faced by a perfectly competitive firm is below average variable cost at the profit-maximizing quantity of output, then the firm should shut down operations immediately. If the market price faced by a perfectly competitive firm is above average variable cost, but below average cost, then the firm should continue producing in the short run, but exit in the long run. The point where the marginal cost curve crosses the average variable cost curve is called the shutdown point.

## Self-Check Questions

### Exercise:

#### Problem:

Look at [\[link\]](#). What would happen to the firm’s profits if the market price increases to \$6 per pack of raspberries?

Quantity	Total Cost	Fixed Cost	Variable Cost	Total Revenue	Profit
----------	------------	------------	---------------	---------------	--------

Quantity	Total Cost	Fixed Cost	Variable Cost	Total Revenue	Profit
0	\$62	\$62	-	\$0	-\$62
10	\$90	\$62	\$28	\$60	-\$30
20	\$110	\$62	\$48	\$120	\$10
30	\$126	\$62	\$64	\$180	\$54
40	\$144	\$62	\$82	\$240	\$96
50	\$166	\$62	\$104	\$300	\$134
60	\$192	\$62	\$130	\$360	\$168
70	\$224	\$62	\$162	\$420	\$196
80	\$264	\$62	\$202	\$480	\$216
90	\$324	\$62	\$262	\$540	\$216
100	\$404	\$62	\$342	\$600	\$196

---

**Solution:**

Holding total cost constant, profits at every output level would increase.

**Exercise:**

**Problem:**

Suppose that the market price increases to \$6, as shown in [\[link\]](#). What would happen to the profit-maximizing output level?

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
0	\$62	\$62	-	-	\$0	-
10	\$90	\$62	\$28	\$2.80	\$60	\$6.00
20	\$110	\$62	\$48	\$2.00	\$120	\$6.00
30	\$126	\$62	\$64	\$1.60	\$180	\$6.00
40	\$144	\$62	\$82	\$1.80	\$240	\$6.00
50	\$166	\$62	\$104	\$2.20	\$300	\$6.00

---

Quantity	Total Cost	Fixed Cost	Variable Cost	Marginal Cost	Total Revenue	Marginal Revenue
60	\$192	\$62	\$130	\$2.60	\$360	\$6.00
70	\$224	\$62	\$162	\$3.20	\$420	\$6.00
80	\$264	\$62	\$202	\$4.00	\$480	\$6.00
<b>90</b>	\$324	\$62	\$262	<b>\$6.00</b>	\$540	<b>\$6.00</b>
100	\$404	\$62	\$342	\$8.00	\$600	\$6.00

---

**Solution:**

When the market price increases, marginal revenue increases. The firm would then increase production up to the point where the new price equals marginal cost, at a quantity of 90.

**Exercise:**

**Problem:**

Explain in words why a profit-maximizing firm will not choose to produce at a quantity where marginal cost exceeds marginal revenue.

---

**Solution:**

If marginal costs exceeds marginal revenue, then the firm will reduce its profits for every additional unit of output it produces. Profit would be greatest if it reduces output to where  $MR = MC$ .

**Exercise:**

**Problem:**

A firm's marginal cost curve above the average variable cost curve is equal to the firm's individual supply curve. This means that every time a firm receives a price from the market it will be willing to supply the amount of output where the price equals marginal cost. What happens to the firm's individual supply curve if marginal costs increase?

---

**Solution:**

The firm will be willing to supply fewer units at every price level. In other words, the firm's individual supply curve decreases and shifts to the left.

## Review Questions

**Exercise:**

**Problem:** How does a perfectly competitive firm decide what price to charge?

**Exercise:**

**Problem:**

What prevents a perfectly competitive firm from seeking higher profits by increasing the price that it charges?

**Exercise:**

**Problem:** How does a perfectly competitive firm calculate total revenue?

**Exercise:**

**Problem:**

Briefly explain the reason for the shape of a marginal revenue curve for a perfectly competitive firm.

**Exercise:**

**Problem:**

What two rules does a perfectly competitive firm apply to determine its profit-maximizing quantity of output?

**Exercise:**

**Problem:** How does the average cost curve help to show whether a firm is making profits or losses?

**Exercise:**

**Problem:** What two lines on a cost curve diagram intersect at the zero-profit point?

**Exercise:**

**Problem:** Should a firm shut down immediately if it is making losses?

**Exercise:**

**Problem:**

How does the average variable cost curve help a firm know whether it should shut down immediately?

**Exercise:**

**Problem:** What two lines on a cost curve diagram intersect at the shutdown point?

## Critical Thinking Questions

**Exercise:**

**Problem:**

Your company operates in a perfectly competitive market. You have been told that advertising can help you increase your sales in the short run. Would you create an aggressive advertising campaign for your product?

**Exercise:**

**Problem:**

Since a perfectly competitive firm can sell as much as it wishes at the market price, why can the firm not simply increase its profits by selling an extremely high quantity?

## Problems

**Exercise:**

**Problem:**

The AAA Aquarium Co. sells aquariums for \$20 each. Fixed costs of production are \$20. The total variable costs are \$20 for one aquarium, \$25 for two units, \$35 for the three units, \$50 for four units, and \$80 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost, and marginal cost for each output level (one to five units). What is the profit-maximizing quantity of output? On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves.

**Exercise:****Problem:**

Perfectly competitive firm Doggies Paradise Inc. sells winter coats for dogs. Dog coats sell for \$72 each. The fixed costs of production are \$100. The total variable costs are \$64 for one unit, \$84 for two units, \$114 for three units, \$184 for four units, and \$270 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost and marginal cost for each output level (one to five units). On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves. What is the profit maximizing quantity?

**Exercise:****Problem:**

A computer company produces affordable, easy-to-use home computer systems and has fixed costs of \$250. The marginal cost of producing computers is \$700 for the first computer, \$250 for the second, \$300 for the third, \$350 for the fourth, \$400 for the fifth, \$450 for the sixth, and \$500 for the seventh.

- Create a table that shows the company's output, total cost, marginal cost, average cost, variable cost, and average variable cost.
- At what price is the zero-profit point? At what price is the shutdown point?
- If the company sells the computers for \$500, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.
- If the firm sells the computers for \$300, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.

**Glossary**

marginal revenue

the additional revenue gained from selling one more unit

shutdown point

level of output where the marginal cost curve intersects the average variable cost curve at the minimum point of AVC; if the price is below this point, the firm should shut down immediately

## Entry and Exit Decisions in the Long Run

By the end of this section, you will be able to:

- Explain how entry and exit lead to zero profits in the long run
- Discuss the long-run adjustment process

The line between the short run and the long run cannot be defined precisely with a stopwatch, or even with a calendar. It varies according to the specific business. The distinction between the short run and the long run is therefore more technical: in the short run, firms cannot change the usage of fixed inputs, while in the long run, the firm can adjust all factors of production.

In a competitive market, profits are a red cape that incites businesses to charge. If a business is making a profit in the short run, it has an incentive to expand existing factories or to build new ones. New firms may start production, as well. When new firms enter the industry in response to increased industry profits it is called **entry**.

Losses are the black thundercloud that causes businesses to flee. If a business is making losses in the short run, it will either keep limping along or just shut down, depending on whether its revenues are covering its variable costs. But in the long run, firms that are facing losses will shut down at least some of their output, and some firms will cease production altogether. The long-run process of reducing production in response to a sustained pattern of losses is called **exit**. The following Clear It Up feature discusses where some of these losses might come from, and the reasons why some firms go out of business.

### **Note:**

Why do firms cease to exist?

Can we say anything about what causes a firm to exit an industry? Profits are the measurement that determines whether a business stays operating or not. Individuals start businesses with the purpose of making profits. They invest their money, time, effort, and many other resources to produce and sell something that they hope will give them something in return.

Unfortunately, not all businesses are successful, and many new startups soon realize that their “business adventure” must eventually end. In the model of perfectly competitive firms, those that consistently cannot make money will “exit,” which is a nice, bloodless word for a more painful process. When a business fails, after all, workers lose their jobs, investors lose their money, and owners and managers can lose their dreams. Many businesses fail. The U.S. Small Business Administration indicates that in 2011, 409,040 new firms “entered,” and 470,376 firms failed. Sometimes a business fails because of poor management or workers who are not very productive, or because of tough domestic or foreign competition. Businesses also fail from a variety of causes that might best be summarized as bad luck. For example, conditions of demand and supply in the market shift in an unexpected way, so that the prices that can be charged for outputs fall or the prices that need to be paid for inputs rise. With millions of businesses in the U.S. economy, even a small fraction of them failing will affect many people—and business failures can be very hard on the workers and managers directly involved. But from the standpoint of the overall economic system, business exits are sometimes a necessary evil if a market-oriented system is going to offer a flexible mechanism for satisfying customers, keeping costs low, and inventing new products.

## **How Entry and Exit Lead to Zero Profits in the Long Run**

No perfectly competitive firm acting alone can affect the market price. However, the combination of many firms entering or exiting the market will affect overall supply in the market. In turn, a shift in supply for the market as a whole will affect the market price. Entry and exit to and from the market are the driving forces behind a process that, in the long run, pushes the price down to minimum average total costs so that all firms are earning a zero profit.

To understand how short-run profits for a perfectly competitive firm will evaporate in the long run, imagine the following situation. The market is in **long-run equilibrium**, where all firms earn zero economic profits

producing the output level where  $P = MR = MC$  and  $P = AC$ . No firm has the incentive to enter or leave the market. Let's say that the product's demand increases, and with that, the market price goes up. The existing firms in the industry are now facing a higher price than before, so they will increase production to the new output level where  $P = MR = MC$ .

This will temporarily make the market price rise above the average cost curve, and therefore, the existing firms in the market will now be earning economic profits. However, these economic profits attract other firms to enter the market. Entry of many new firms causes the market supply curve to shift to the right. As the supply curve shifts to the right, the market price starts decreasing, and with that, economic profits fall for new and existing firms. As long as there are still profits in the market, entry will continue to shift supply to the right. This will stop whenever the market price is driven down to the zero-profit level, where no firm is earning economic profits.

Short-run losses will fade away by reversing this process. Say that the market is in long-run equilibrium. This time, instead, demand decreases, and with that, the market price starts falling. The existing firms in the industry are now facing a lower price than before, and as it will be below the average cost curve, they will now be making economic losses. Some firms will continue producing where the new  $P = MR = MC$ , as long as they are able to cover their average variable costs. Some firms will have to shut down immediately as they will not be able to cover their average variable costs, and will then only incur their fixed costs, minimizing their losses. Exit of many firms causes the market supply curve to shift to the left. As the supply curve shifts to the left, the market price starts rising, and economic losses start to be lower. This process ends whenever the market price rises to the zero-profit level, where the existing firms are no longer losing money and are at zero profits again. Thus, while a perfectly competitive firm can earn profits in the short run, in the long run the process of entry will push down prices until they reach the zero-profit level. Conversely, while a perfectly competitive firm may earn losses in the short run, firms will not continually lose money. In the long run, firms making losses are able to escape from their fixed costs, and their exit from the market will push the price back up to the zero-profit level. In the long run, this process of entry and exit will drive the price in perfectly competitive

markets to the zero-profit point at the bottom of the AC curve, where marginal cost crosses average cost.

## **The Long-Run Adjustment and Industry Types**

Whenever there are expansions in an industry, costs of production for the existing and new firms could either stay the same, increase, or even decrease. Therefore, we can categorize an industry as being (1) a constant cost industry (as demand increases, the cost of production for firms stays the same), (2) an increasing cost industry (as demand increases, the cost of production for firms increases), or (3) a decreasing cost industry (as demand increases the costs of production for the firms decreases).

For a constant cost industry, whenever there is an increase in market demand and price, then the supply curve shifts to the right with new firms' entry and stops at the point where the new long-run equilibrium intersects at the same market price as before. But why will costs remain the same? In this type of industry, the supply curve is very elastic. Firms can easily supply any quantity that consumers demand. In addition, there is a perfectly elastic supply of inputs—firms can easily increase their demand for employees, for example, with no increase to wages. Tying in to our Bring it Home discussion, an increased demand for ethanol in recent years has caused the demand for corn to increase. Consequently, many farmers switched from growing wheat to growing corn. Agricultural markets are generally good examples of constant cost industries.

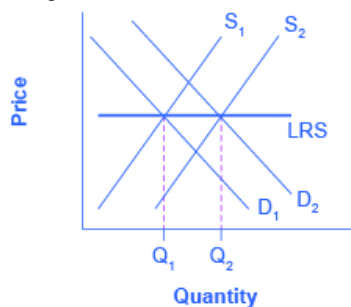
For an increasing cost industry, as the market expands, the old and new firms experience increases in their costs of production, which makes the new zero-profit level intersect at a higher price than before. Here companies may have to deal with limited inputs, such as skilled labor. As the demand for these workers rise, wages rise and this increases the cost of production for all firms. The industry supply curve in this type of industry is more inelastic.

For a decreasing cost industry, as the market expands, the old and new firms experience lower costs of production, which makes the new zero-profit level intersect at a lower price than before. In this case, the industry

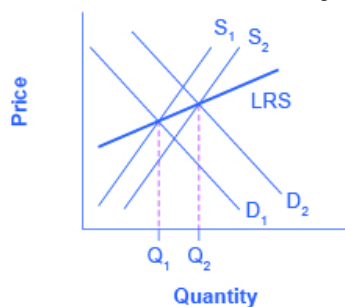
and all the firms in it are experiencing falling average total costs. This can be due to an improvement in technology in the entire industry or an increase in the education of employees. High tech industries may be a good example of a decreasing cost market.

[\[link\]](#) (a) presents the case of an adjustment process in a constant cost industry. Whenever there are output expansions in this type of industry, the long-run outcome implies more output produced at exactly the same original price. Note that supply was able to increase to meet the increased demand. When we join the before and after long-run equilibriums, the resulting line is the long run supply (LRS) curve in perfectly competitive markets. In this case, it is a flat curve. [\[link\]](#) (b) and [\[link\]](#) (c) present the cases for an increasing cost and decreasing cost industry, respectively. For an increasing cost industry, the LRS is upward sloping, while for a decreasing cost industry, the LRS is downward sloping.

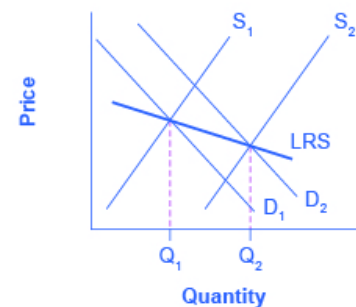
#### Adjustment Process in a Constant-Cost Industry



(a) Constant cost



(b) Increasing cost



(c) Decreasing cost

In (a), demand increased and supply met it. Notice that the supply increase is equal to the demand increase. The result is that the equilibrium price stays the same as quantity sold increases. In (b), notice that sellers were not able to increase supply as much as demand. Some inputs were scarce, or wages were rising. The equilibrium price rises. In (c), sellers easily increased supply in response to the demand increase. Here, new technology or economies of scale caused the large increase in supply, resulting in declining equilibrium price.

## Key Concepts and Summary

In the long run, firms will respond to profits through a process of entry, where existing firms expand output and new firms enter the market. Conversely, firms will react to losses in the long run through a process of exit, in which existing firms reduce output or cease production altogether. Through the process of entry in response to profits and exit in response to losses, the price level in a perfectly competitive market will move toward the zero-profit point, where the marginal cost curve crosses the AC curve, at the minimum of the average cost curve.

The long-run supply curve shows the long-run output supplied by firms in three different types of industries: constant cost, increasing cost, and decreasing cost.

## Self-Check Questions

### Exercise:

#### Problem:

If new technology in a perfectly competitive market brings about a substantial reduction in costs of production, how will this affect the market?

---

#### Solution:

With a technological improvement that brings about a reduction in costs of production, an adjustment process will take place in the market. The technological improvement will result in an increase in supply curves, by individual firms and at the market level. The existing firms will experience higher profits for a while, which will attract other firms into the market. This entry process will stop whenever the market supply increases enough (both by existing and new firms) so profits are driven back to zero.

### Exercise:

**Problem:**

A market in perfect competition is in long-run equilibrium. What happens to the market if labor unions are able to increase wages for workers?

---

**Solution:**

When wages increase, costs of production increase. Some firms would now be making economic losses and would shut down. The supply curve then starts shifting to the left, pushing the market price up. This process ends when all firms remaining in the market earn zero economic profits. The result is a contraction in the output produced in the market.

**Review Questions****Exercise:**

**Problem:** Why does entry occur?

**Exercise:**

**Problem:** Why does exit occur?

**Exercise:****Problem:**

Do entry and exit occur in the short run, the long run, both, or neither?

**Exercise:****Problem:**

What price will a perfectly competitive firm end up charging in the long run? Why?

## Critical Thinking Questions

### Exercise:

#### Problem:

Many firms in the United States file for bankruptcy every year, yet they still continue operating. Why would they do this instead of completely shutting down?

### Exercise:

#### Problem:

Why will profits for firms in a perfectly competitive industry tend to vanish in the long run?

### Exercise:

#### Problem:

Why will losses for firms in a perfectly competitive industry tend to vanish in the long run?

## Glossary

### entry

the long-run process of firms entering an industry in response to industry profits

### exit

the long-run process of firms reducing production and shutting down in response to industry losses

### long-run equilibrium

where all firms earn zero economic profits producing the output level where  $P = MR = MC$  and  $P = AC$

## Efficiency in Perfectly Competitive Markets

By the end of this section, you will be able to:

- Apply concepts of productive efficiency and allocative efficiency to perfectly competitive markets
- Compare the model of perfect competition to real-world markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that were first introduced in [Choice in a World of Scarcity](#)) .

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, goods are being produced and sold at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the point that is chosen is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that is paid for a good as a measure of the social benefit received for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits received from producing a good are in line with the social costs of production.

To explore what is meant by allocative efficiency, it is useful to walk through an example. Begin by assuming that the market for wholesale flowers is perfectly competitive, and so  $P = MC$ . Now, consider what it would mean if firms in that market produced a lesser quantity of flowers. At a lesser quantity, marginal costs will not yet have increased as much, so that

price will exceed marginal cost; that is,  $P > MC$ . In that situation, the benefit to society as a whole of producing additional goods, as measured by the willingness of consumers to pay for marginal units of a good, would be higher than the cost of the inputs of labor and physical capital needed to produce the marginal good. In other words, the gains to society as a whole from producing additional marginal units will be greater than the costs.

Conversely, consider what it would mean if, compared to the level of output at the allocatively efficient choice when  $P = MC$ , firms produced a greater quantity of flowers. At a greater quantity, marginal costs of production will have increased so that  $P < MC$ . In that case, the marginal costs of producing additional flowers is greater than the benefit to society as measured by what people are willing to pay. For society as a whole, since the costs are outstripping the benefits, it will make sense to produce a lower quantity of such goods.

When perfectly competitive firms maximize their profits by producing the quantity where  $P = MC$ , they also assure that the benefits to consumers of what they are buying, as measured by the price they are willing to pay, is equal to the costs to society of producing the marginal units, as measured by the marginal costs the firm must pay—and thus that allocative efficiency holds.

The statements that a perfectly competitive market in the long run will feature both productive and allocative efficiency do need to be taken with a few grains of salt. Remember, economists are using the concept of “efficiency” in a particular and specific sense, not as a synonym for “desirable in every way.” For one thing, consumers’ ability to pay reflects the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because they have insufficient income.

Perfect competition, in the long run, is a hypothetical benchmark. For market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many issues that are assumed away in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defense or education, discrimination in labor markets, and buyers and sellers who must deal with imperfect and unclear information. These issues are explored in other chapters. However, the theoretical efficiency of perfect competition does provide a useful benchmark for comparing the issues that arise from these real-world problems.

**Note:**

**A Dime a Dozen**

A quick glance at [\[link\]](#) reveals the dramatic increase in North Dakota corn production—more than double. Taking into consideration that corn typically yields two to three times as many bushels per acre as wheat, it is obvious there has been a significant increase in bushels of corn. Why the increase in corn acreage? Converging prices.

Year	Corn (millions of acres)	Wheat (millions of acres)
2014	91.6	56.82

(Source: USDA National Agricultural Statistics Service)

Historically, wheat prices have been higher than corn prices, offsetting wheat's lower yield per acre. However, in recent years wheat and corn prices have been converging. In April 2013, *Agweek* reported the gap was just 71 cents per bushel. As the difference in price narrowed, switching to the production of higher yield per acre of corn simply made good business sense. Erik Younggren, president of the National Association of Wheat Growers said in the *Agweek* article, "I don't think we're going to see mile

after mile of waving amber fields [of wheat] anymore." (Until wheat prices rise, we will probably be seeing field after field of tasseled corn.)

## Key Concepts and Summary

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

## Self-Check Questions

### Exercise:

#### Problem:

Productive efficiency and allocative efficiency are two concepts achieved in the long run in a perfectly competitive market. These are the two reasons why we call them “perfect.” How would you use these two concepts to analyze other market structures and label them “imperfect?”

---

#### Solution:

Perfect competition is considered to be “perfect” because both allocative and productive efficiency are met at the same time in a long-run equilibrium. If a market structure results in long-run equilibrium that does not minimize average total costs and/or does not charge a price equal to marginal cost, then either allocative or productive (or both) efficiencies are not met, and therefore the market cannot be labeled “perfect.”

### Exercise:

**Problem:**

Explain how the profit-maximizing rule of setting  $P = MC$  leads a perfectly competitive market to be allocatively efficient.

---

**Solution:**

Think of the market price as representing the gain to society from a purchase, since it represents what someone is willing to pay. Think of the marginal cost as representing the cost to society from making the last unit of a good. If  $P > MC$ , then the benefits from producing more of a good exceed the costs, and society would gain from producing more of the good. If  $P < MC$ , then the social costs of producing the marginal good exceed the social benefits, and society should produce less of the good. Only if  $P = MC$ , the rule applied by a profit-maximizing perfectly competitive firm, will society's costs and benefits be in balance. This choice will be the option that brings the greatest overall benefit to society.

**Review Questions****Exercise:****Problem:**

Will a perfectly competitive market display productive efficiency? Why or why not?

**Exercise:****Problem:**

Will a perfectly competitive market display allocative efficiency? Why or why not?

**Critical Thinking Questions**

**Exercise:****Problem:**

Assuming that the market for cigarettes is in perfect competition, what does allocative and productive efficiency imply in this case? What does it not imply?

**Exercise:****Problem:**

In the argument for why perfect competition is allocatively efficient, the price that people are willing to pay represents the gains to society and the marginal cost to the firm represents the costs to society. Can you think of some social costs or issues that are not included in the marginal cost to the firm? Or some social gains that are not included in what people pay for a good?

**References**

Index Mundi. n.d. "Wheat Monthly Price—U.S. Dollars per Metric Ton." Accessed March 11, 2015. <http://www.indexmundi.com/commodities/?commodity=wheat>.

Knutson, J. "Wheat on the Defensive in the Northern Plains." *Agweek, Associated Press State Wire: North Dakota (ND)*. April 14, 2013.

SBA Office of Advocacy. 2014. "Frequently Asked Questions: Advocacy: the voice of small business in government." Accessed March 11, 2015. [https://www.sba.gov/sites/default/files/advocacy/FAQ\\_March\\_2014\\_0.pdf](https://www.sba.gov/sites/default/files/advocacy/FAQ_March_2014_0.pdf).

## Introduction to a Monopoly

class="introduction"

### Political Power from a Cotton Monopoly

In the mid-nineteenth century, the United States, specifically the Southern states, had a near monopoly in the cotton supplied to Great Britain. These states attempted to leverage this economic power into political power—trying to sway Great Britain to formally recognize the Confederate States of America. (Credit: modification of work by “ashleylovespizza”/Flickr or Creative Commons)

**Note:****The Rest is History**

Many of the opening case studies have focused on current events. This one steps into the past to observe how monopoly, or near monopolies, have helped shape history. In the spring of 1773, the East India Company, a firm that, in its time, was designated ‘too big to fail,’ was continuing to experience financial difficulties. To help shore up the failing firm, the British Parliament authorized the Tea Act. The act continued the tax on teas and made the East India Company the sole legal supplier of tea to the American colonies. By November, the citizens of Boston had had enough. They refused to permit the tea to be unloaded, citing their main complaint: “No taxation without representation.” Arriving tea-bearing ships were warned via several newspapers, including *The Massachusetts Gazette*, “We are prepared, and shall not fail to pay them an unwelcome visit; by The Mohawks.”

Step forward in time to 1860—the eve of the American Civil War—to another near monopoly supplier of historical significance: the U.S. cotton industry. At that time, the Southern states provided the majority of the

cotton Britain imported. The South, wanting to secede from the Union, hoped to leverage Britain's high dependency on its cotton into formal diplomatic recognition of the Confederate States of America.

This leads us to the topic of this chapter: a firm that controls all (or nearly all) of the supply of a good or service—a monopoly. How do monopoly firms behave in the marketplace? Do they have “power?” Does this power potentially have unintended consequences? We'll return to this case at the end of the chapter to see how the tea and cotton monopolies influenced U.S. history.

**Note:**

**Introduction to a Monopoly**

In this chapter, you will learn about:

- How Monopolies form: Barriers to Entry
- How a Profit-Maximizing Monopoly Chooses Output and Price

There is a widespread belief that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you are? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the chapter on [Perfect Competition](#). In this chapter, we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have complete market power. In the case of **monopoly**, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it

wishes. While a monopoly, by definition, refers to a single firm, in practice the term is often used to describe a market in which one firm merely has a very high market share. This tends to be the definition that the U.S. Department of Justice uses.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: The U.S. Postal Service, your electric and garbage collection companies are a few examples. Some new drugs are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the U.S. Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department's argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers. In 2015, a U.S. federal court tossed out antitrust charges that Google had an agreement with mobile device makers to set Google as the default search engine.

This chapter begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain configurations of demand and supply. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

## How Monopolies Form: Barriers to Entry

By the end of this section, you will be able to:

- Distinguish between a natural monopoly and a legal monopoly.
- Explain how economies of scale and the control of natural resources led to the necessary formation of legal monopolies
- Analyze the importance of trademarks and patents in promoting innovation
- Identify examples of predatory pricing

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract vigorous competition as described in [Perfect Competition](#), and yet, because of one particular characteristic of monopoly, they do not. **Barriers to entry** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable, such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once the rights to all of them have been purchased, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

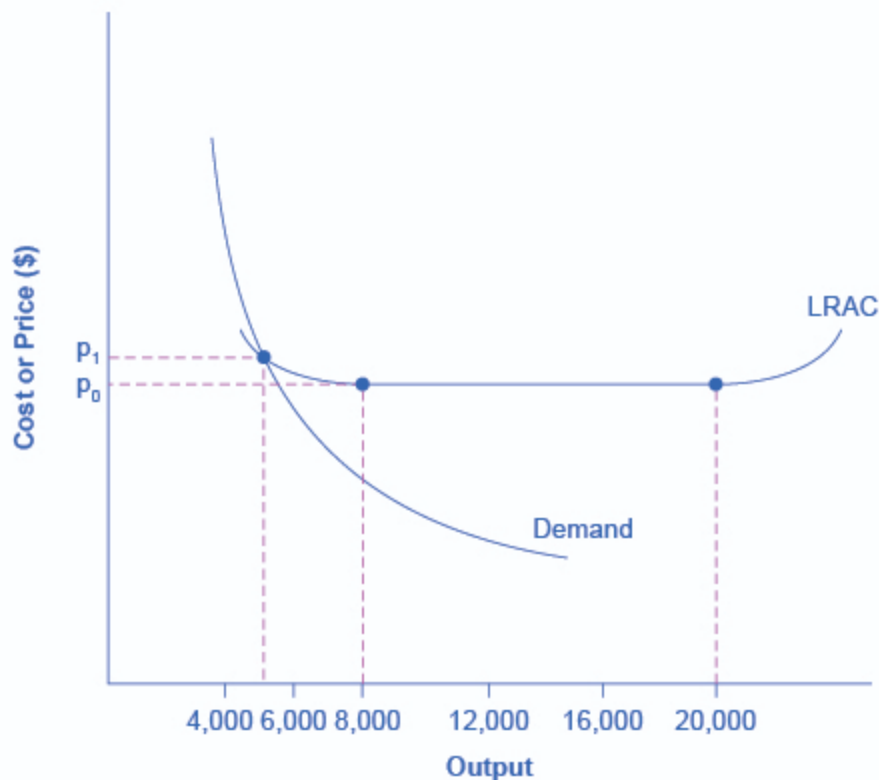
There are two types of monopoly, based on the types of barriers to entry they exploit. One is **natural monopoly**, where the barriers to entry are something other than legal prohibition. The other is **legal monopoly**, where laws prohibit (or severely limit) competition.

### Natural Monopoly

Economies of scale can combine with the size of the market to limit competition. (This theme was introduced in [Cost and Industry Structure](#)). [\[link\]](#) presents a long-run average cost curve for the airplane manufacturing industry. It shows economies of scale up to an output of 8,000 planes per year and a price of  $P_0$ , then constant returns to scale from 8,000 to 20,000 planes per year, and diseconomies of scale at a quantity of production greater than 20,000 planes per year.

Now consider the market demand curve in the diagram, which intersects the long-run average cost (LRAC) curve at an output level of 6,000 planes per year and at a price  $P_1$ , which is higher than  $P_0$ . In this situation, the market has room for only one producer. If a second firm attempts to enter the market at a smaller size, say by producing a quantity of 4,000 planes, then its average costs will be higher than the existing firm, and it will be unable to compete. If the second firm attempts to enter the market at a larger size, like 8,000 planes per year, then it could produce at a lower average cost—but it could not sell all 8,000 planes that it produced because of insufficient demand in the market.

Economies of Scale and Natural Monopoly



In this market, the demand curve intersects the long-run average cost (LRAC) curve at its downward-sloping part. A natural monopoly occurs when the quantity demanded is less than the minimum quantity it takes to be at the bottom of the long-run average cost curve.

This situation, when economies of scale are large relative to the quantity demanded in the market, is called a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. Once the main water pipes are laid through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once electricity lines are installed through a neighborhood, the marginal cost of providing additional electrical service to one more home is very low. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport. For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

## **Control of a Physical Resource**

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the U.S. economy, one historical example of this pattern occurred when ALCOA—the Aluminum Company of America—controlled most of the supply of bauxite, a key mineral used in making

aluminum. Back in the 1930s, when ALCOA controlled most of the bauxite, other firms were simply unable to produce enough aluminum to compete.

As another example, the majority of global diamond production is controlled by DeBeers, a multi-national company that has mining and production operations in South Africa, Botswana, Namibia, and Canada. It also has exploration activities on four continents, while directing a worldwide distribution network of rough cut diamonds. Though in recent years they have experienced growing competition, their impact on the rough diamond market is still considerable.

## **Legal Monopoly**

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under U.S. law, no organization but the U.S. Postal Service is legally allowed to deliver first-class mail. Many states or cities have laws or regulations that allow households a choice of only one electric company, one water company, and one company to pick up the garbage. Most legal monopolies are considered utilities—products necessary for everyday life—that are socially beneficial to have. As a consequence, the government allows producers to become regulated monopolies, to insure that an appropriate amount of these products is provided to consumers. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

## **Promoting Innovation**

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In this world of near ubiquitous information, other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R&D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation. To prevent this from happening, the Constitution of the United States specifies in Article I, Section 8: “The Congress shall have

Power . . . To Promote the Progress of Science and Useful Arts, by securing for limited Times to Authors and Inventors the Exclusive Right to their Writings and Discoveries.” Congress used this power to create the U.S. Patent and Trademark Office, as well as the U.S. Copyright Office. A **patent** gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time; in the United States, exclusive patent rights last for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in R&D, but then to allow other firms to produce the product more cheaply once the patent expires.

A **trademark** is an identifying symbol or name for a particular good, like Chiquita bananas, Chevrolet cars, or the Nike “swoosh” that appears on shoes and athletic gear. Roughly 1.9 million trademarks are registered with the U.S. government. A firm can renew a trademark over and over again, as long as it remains in active use.

A **copyright**, according to the U.S. Copyright Office, “is a form of protection provided by the laws of the United States for ‘original works of authorship’ including literary, dramatic, musical, architectural, cartographic, choreographic, pantomimic, pictorial, graphic, sculptural, and audiovisual creations.” No one can reproduce, display, or perform a copyrighted work without permission of the author. Copyright protection ordinarily lasts for the life of the author plus 70 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. But in certain areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as **trade secrets**. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, this combination of patents, trademarks, copyrights, and trade secret law is called **intellectual property**, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions

of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony to the intellectual property laws of different countries to determine the extent to which patents and copyrights in one country will be respected in other countries.

Government limitations on competition used to be even more common in the United States. For most of the twentieth century, only one phone company—AT&T—was legally allowed to provide local and long distance service. From the 1930s to the 1970s, one set of federal regulations limited which destinations airlines could choose to fly to and what fares they could charge; another set of regulations limited the interest rates that banks could pay to depositors; yet another specified what trucking firms could charge customers.

What products are considered utilities depends, in part, on the available technology. Fifty years ago, local and long distance telephone service was provided over wires. It did not make much sense to have multiple companies building multiple systems of wiring across towns and across the country. AT&T lost its monopoly on long distance service when the technology for providing phone service changed from wires to microwave and satellite transmission, so that multiple firms could use the same transmission mechanism. The same thing happened to local service, especially in recent years, with the growth in cellular phone systems.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of **deregulation**, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that could be charged, and the quantities that could be produced in many industries, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

**Note:**

Visit this [website](#) for examples of some pretty bizarre patents.



## Intimidating Potential Competition

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as **predatory pricing**, in which a firm uses the threat of sharp price cuts to discourage competition. Predatory pricing is a violation of U.S. antitrust law, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the incumbent firm can raise prices again.

After this pattern is repeated once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse larger airlines of predatory pricing: in the early 2000s, for example, ValuJet accused Delta of predatory pricing, Frontier accused United, and Reno Air accused Northwest. In 2015, the Justice Department ruled against American Express and Mastercard for imposing restrictions on retailers who encouraged customers to use lower swipe fees on credit transactions.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new

national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

## Summing Up Barriers to Entry

[\[link\]](#) lists the barriers to entry that have been discussed here. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Barrier to Entry	Government Role?	Example
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and trucking
Patent, trademark, and copyright	Yes, through protection of intellectual property	New drugs or software

<b>Barrier to Entry</b>	<b>Government Role?</b>	<b>Example</b>
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names

Barriers to Entry

## Key Concepts and Summary

Barriers to entry prevent or discourage competitors from entering the market. These barriers include: economies of scale that lead to natural monopoly; control of a physical resource; legal restrictions on competition; patent, trademark and copyright protection; and practices to intimidate the competition like predatory pricing. Intellectual property refers to legally guaranteed ownership of an idea, rather than a physical item. The laws that protect intellectual property include patents, copyrights, trademarks, and trade secrets. A natural monopoly arises when economies of scale persist over a large enough range of output that if one firm supplies the entire market, no other firm can enter without facing a cost disadvantage.

## Self-Check Questions

### Exercise:

#### Problem:

Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

- A patented invention
- A popular but easily copied restaurant recipe
- An industry where economies of scale are very small compared to the size of demand in the market

- d. A well-established reputation for slashing prices in response to new entry
  - e. A well-respected brand name that has been carefully built up over many years
- 

**Solution:**

- a. A patent is a government-enforced barrier to entry.
- b. This is not a barrier to entry.
- c. This is not a barrier to entry.
- d. This is a barrier to entry, but it is not government-enforced.
- e. This is a barrier to entry, but it is not directly government enforced.

**Exercise:**

**Problem:**

Classify the following as a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or a situation that does not involve a barrier to entry.

- a. A city passes a law on how many licenses it will issue for taxicabs
  - b. A city passes a law that all taxicab drivers must pass a driving safety test and have insurance
  - c. A well-known trademark
  - d. Owning a spring that offers very pure water
  - e. An industry where economies of scale are very large compared to the size of demand in the market
- 

**Solution:**

- a. This is a government-enforced barrier to entry.
- b. This is an example of a government law, but perhaps it is not much of a barrier to entry if most people can pass the safety test

- and get insurance.
- c. Trademarks are enforced by government, and therefore are a barrier to entry.
  - d. This is probably not a barrier to entry, since there are a number of different ways of getting pure water.
  - e. This is a barrier to entry, but it is not government-enforced.

**Exercise:**

**Problem:**

Suppose the local electrical utility, a legal monopoly based on economies of scale, was split into four firms of equal size, with the idea that eliminating the monopoly would promote competitive pricing of electricity. What do you anticipate would happen to prices?

---

**Solution:**

Because of economies of scale, each firm would produce at a higher average cost than before. (They would each have to build their own power lines.) As a result, they would each have to raise prices to cover their higher costs. The policy would fail.

**Exercise:**

**Problem:**

If Congress reduced the period of patent protection from 20 years to 10 years, what would likely happen to the amount of private research and development?

---

**Solution:**

Shorter patent protection would make innovation less lucrative, so the amount of research and development would likely decline.

**Review Questions**

**Exercise:**

**Problem:** How is monopoly different from perfect competition?

**Exercise:**

**Problem:** What is a barrier to entry? Give some examples.

**Exercise:**

**Problem:** What is a natural monopoly?

**Exercise:**

**Problem:** What is a legal monopoly?

**Exercise:**

**Problem:** What is predatory pricing?

**Exercise:**

**Problem:** How is intellectual property different from other property?

**Exercise:**

**Problem:**

By what legal mechanisms is intellectual property protected?

**Exercise:**

**Problem:** In what sense is a natural monopoly “natural”?

## **Critical Thinking Questions**

**Exercise:**

**Problem:**

ALCOA does not have the monopoly power it once had. How do you suppose their barriers to entry were weakened?

**Exercise:****Problem:**

Why are generic pharmaceuticals significantly cheaper than name brand ones?

**Exercise:****Problem:**

For many years, the Justice Department has tried to break up large firms like IBM, Microsoft, and most recently Google, on the grounds that their large market share made them essentially monopolies. In a global market, where U.S. firms compete with firms from other countries, would this policy make the same sense as it might in a purely domestic context?

**Exercise:****Problem:**

Intellectual property laws are intended to promote innovation, but some economists, such as Milton Friedman, have argued that such laws are not desirable. In the United States, there is no intellectual property protection for food recipes or for fashion designs. Considering the state of these two industries, and bearing in mind the discussion of the inefficiency of monopolies, can you think of any reasons why intellectual property laws might hinder innovation in some cases?

**Problems****Exercise:**

**Problem:**

Return to [\[link\]](#). Suppose  $P_0$  is \$10 and  $P_1$  is \$11. Suppose a new firm with the same LRAC curve as the incumbent tries to break into the market by selling 4,000 units of output. Estimate from the graph what the new firm's average cost of producing output would be. If the incumbent continues to produce 6,000 units, how much output would be supplied to the market by the two firms? Estimate what would happen to the market price as a result of the supply of both the incumbent firm and the new entrant. Approximately how much profit would each firm earn?

**Glossary**

barriers to entry

the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright

a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation

removing government controls over setting prices and quantities in certain industries

intellectual property

the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their inventions

legal monopoly

legal prohibitions against competition, such as regulated monopolies and intellectual property protection

monopoly

a situation in which one firm produces all of the output in a market

natural monopoly

economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent

a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing

when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets

methods of production kept secret by the producing firm

trademark

an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

## How a Profit-Maximizing Monopoly Chooses Output and Price

By the end of this section, you will be able to:

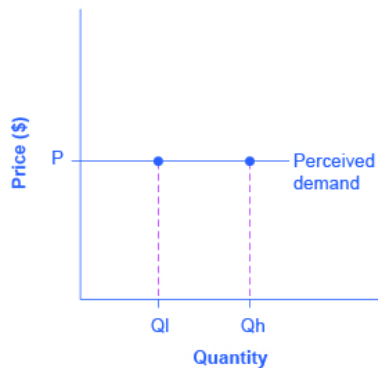
- Explain the perceived demand curve for a perfect competitor and a monopoly
- Analyze a demand curve for a monopoly and determine the output that maximizes profit and revenue
- Calculate marginal revenue and marginal cost
- Explain allocative efficiency as it pertains to the efficiency of a monopoly

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. The pattern of costs for the monopoly can be analyzed within the same framework as the costs of a perfectly competitive firm—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm. (The Clear it Up feature discusses how hard it is sometimes to define “market” in a monopoly situation.)

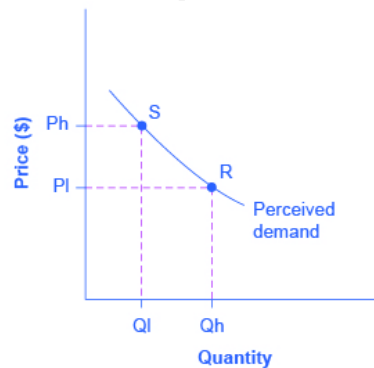
### Demand Curves Perceived by a Perfectly Competitive Firm and by a Monopoly

A perfectly competitive firm acts as a price taker, so its calculation of total revenue is made by taking the given market price and multiplying it by the quantity of output that the firm chooses. The demand curve *as it is perceived by a perfectly competitive firm* appears in [\[link\]](#) (a). The flat perceived demand curve means that, from the viewpoint of the perfectly competitive firm, it could sell either a relatively low quantity like  $Q_l$  or a relatively high quantity like  $Q_h$  at the market price  $P$ .

The Perceived Demand Curve for a Perfect Competitor and a Monopolist



(a) Perceived demand for a perfect competitor



(b) Perceived demand for a monopolist

(a) A perfectly competitive firm perceives the demand curve that it faces to be flat. The flat shape means that the firm can sell either a low quantity ( $Q_l$ ) or a high quantity ( $Q_h$ ) at exactly the same price ( $P$ ). (b) A monopolist perceives the demand curve that it faces to be the same as the market demand curve, which for most goods is downward-sloping. Thus, if the monopolist chooses a high level of output ( $Q_h$ ), it can charge only a relatively low price ( $P_l$ ); conversely, if the monopolist chooses a low level of output ( $Q_l$ ), it can then charge a higher price ( $P_h$ ). The challenge for the monopolist is to choose the combination of price and quantity that maximizes profits.

**Note:****What defines the market?**

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. But what defines the “market”?

In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the “flexible packaging materials,” which includes all other moisture-proof papers, films, and foils. In 1956, after years of legal appeals, the U.S. Supreme Court held that the broader market definition was more appropriate, and the case against DuPont was dismissed.

Questions over how to define the market continue today. True, Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014.

The Greyhound bus company may have a near-monopoly on the market for intercity bus transportation, but it is only a small share of the market for intercity transportation if that market includes private cars, airplanes, and railroad service. DeBeers has a monopoly in diamonds, but it is a much smaller share of the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town in the country may have only one gas station: is this gas station a “monopoly,” or does it compete with gas stations that might be five, 10, or 50 miles away?

In general, if a firm produces a product without close substitutes, then the firm can be considered a monopoly producer in a single market. But if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

While a monopolist can charge *any* price for its product, that price is nonetheless constrained by demand for the firm’s product. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping.

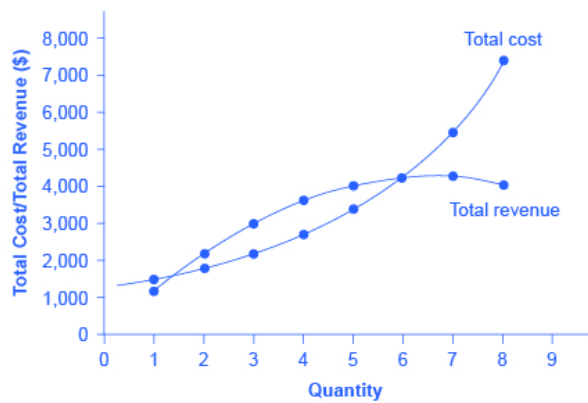
[\[link\]](#) illustrates this situation. The monopolist can either choose a point like R with a low price ( $P_L$ ) and high quantity ( $Q_H$ ), or a point like S with a high price ( $P_H$ ) and a low quantity ( $Q_L$ ), or some intermediate point. Setting the price too high will result in a low quantity sold, and will not bring in much revenue. Conversely, setting the price too low may result in a high quantity sold, but because of the low price, it will not bring in much revenue either. The challenge for the monopolist is to strike a profit-maximizing balance between the price it charges and the quantity that it sells. But why isn’t the perfectly competitive firm’s demand curve also the market demand curve? See the following Clear it Up feature for the answer to this question.

**Note:****What is the difference between perceived demand and market demand?**

The demand curve as perceived by a perfectly competitive firm is not the overall market demand curve for that product. However, the firm’s demand curve as perceived by a monopoly is the same as the market demand curve. The reason for the difference is that each perfectly competitive firm perceives the demand for its products in a market that includes many other firms; in effect, the demand curve perceived by a perfectly competitive firm is a tiny slice of the entire market demand curve. In contrast, a monopoly perceives demand for its product in a market where the monopoly is the only producer.

**Total Cost and Total Revenue for a Monopolist**

Profits for a monopolist can be illustrated with a graph of total revenues and total costs, as shown with the example of the hypothetical HealthPill firm in [link](#). The total cost curve has its typical shape; that is, total costs rise and the curve grows steeper as output increases. Total Revenue and Total Cost for the HealthPill Monopoly



Total revenue for the monopoly firm called HealthPill first rises, then falls. Low levels of output bring in relatively little total revenue, because the quantity is low. High levels of output bring in relatively less revenue, because the high quantity pushes down the market price. The total cost curve is upward-sloping. Profits will be highest at the quantity of output where total revenue is most above total cost. Of the choices in [link](#), the highest profits happen at an output of 4. The profit-maximizing level of output is not the same as the revenue-maximizing level of output, which should make sense, because profits take costs into account and revenues do not.

Quantity	Total Cost	Quantity	Price	Total Revenue	Profit = Total Revenue – Total Cost
1	1,500	1	1,200	1,200	–300
2	1,800	2	1,100	2,200	400
3	2,200	3	1,000	3,000	800
4	2,800	4	900	3,600	800
5	3,500	5	800	4,000	500
6	4,200	6	700	4,200	0
7	5,600	7	600	4,200	–1,400

Quantity	Total Cost	Quantity	Price	Total Revenue	Profit = Total Revenue – Total Cost
8	7,400	8	500	4,000	–3,400

#### Total Costs and Total Revenues of HealthPill

To calculate total revenue for a monopolist, start with the demand curve perceived by the monopolist. [\[link\]](#) shows quantities along the demand curve and the price at each quantity demanded, and then calculates total revenue by multiplying price times quantity at each level of output. (In this example, the output is given as 1, 2, 3, 4, and so on, for the sake of simplicity. If you prefer a dash of greater realism, you can imagine that these output levels and the corresponding prices are measured per 1,000 or 10,000 pills.) As the figure illustrates, total revenue for a monopolist rises, flattens out, and then falls. In this example, total revenue is highest at a quantity of 6 or 7.

Clearly, the total revenue for a monopolist is not a straight upward-sloping line, in the way that total revenue was for a perfectly competitive firm. The different total revenue pattern for a monopolist occurs because the quantity that a monopolist chooses to produce affects the market price, which was not true for a perfectly competitive firm. If the monopolist charges a very high price, then quantity demanded drops, and so total revenue is very low. If the monopolist charges a very low price, then, even if quantity demanded is very high, total revenue will not add up to much. At some intermediate level, total revenue will be highest.

However, the monopolist is not seeking to maximize revenue, but instead to earn the highest possible profit. Profits are calculated in the final row of the table. In the HealthPill example in [\[link\]](#), the highest profit will occur at the quantity where total revenue is the farthest above total cost. Of the choices given in the table, the highest profits occur at an output of 4, where profit is 800.

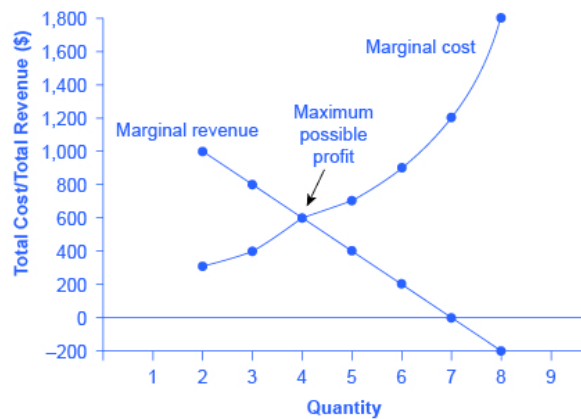
#### Marginal Revenue and Marginal Cost for a Monopolist

In the real world, a monopolist often does not have enough information to analyze its entire total revenues or total costs curves; after all, the firm does not know exactly what would happen if it were to alter production dramatically. But a monopolist often has fairly reliable information about how changing output by small or moderate amounts will affect its marginal revenues and marginal costs, because it has had experience with such changes over time and because modest changes are easier to extrapolate from current experience. A monopolist can use information on marginal revenue and marginal cost to seek out the profit-maximizing combination of quantity and price.

The first four columns of [\[link\]](#) use the numbers on total cost from the HealthPill example in the previous exhibit and calculate marginal cost and average cost. This monopoly faces a typical upward-sloping marginal cost curve, as shown in [\[link\]](#). The second four columns of [\[link\]](#) use the total revenue information from the previous exhibit and calculate marginal revenue.

Notice that marginal revenue is zero at a quantity of 7, and turns negative at quantities higher than 7. It may seem counterintuitive that marginal revenue could ever be zero or negative: after all, does an increase in quantity sold not always mean more revenue? For a perfect competitor, each additional unit sold brought a positive marginal revenue, because marginal revenue was equal to the given market price. But a monopolist can sell a larger quantity and see a decline in total revenue. When a monopolist increases sales by one unit, it gains some marginal revenue from selling that extra unit, but also loses some marginal revenue because every other unit must now be sold at a lower price. As the quantity sold becomes higher, the drop in price affects a greater quantity of sales, eventually causing a situation where more sales cause marginal revenue to be negative.

#### Marginal Revenue and Marginal Cost for the HealthPill Monopoly



For a monopoly like HealthPill, marginal revenue decreases as additional units are sold. The marginal cost curve is upward-sloping. The profit-maximizing choice for the monopoly will be to produce at the quantity where marginal revenue is equal to marginal cost: that is,  $MR = MC$ . If the monopoly produces a lower quantity, then  $MR > MC$  at those levels of output, and the firm can make higher profits by expanding output. If the firm produces at a greater quantity, then  $MC > MR$ , and the firm can make higher profits by reducing its quantity of output.

Cost Information				Revenue Information			
Quantity	Total Cost	Marginal Cost	Average Cost	Quantity	Price	Total Revenue	Marginal Revenue
1	1,500	1,500	1,500	1	1,200	1,200	1,200
2	1,800	300	900	2	1,100	2,200	1,000
3	2,200	400	733	3	1,000	3,000	800
4	2,800	600	700	4	900	3,600	600
5	3,500	700	700	5	800	4,000	400
6	4,200	700	700	6	700	4,200	200
7	5,600	1,400	800	7	600	4,200	0
8	7,400	1,800	925	8	500	4,000	-200

## Costs and Revenues of HealthPill

A monopolist can determine its profit-maximizing price and quantity by analyzing the marginal revenue and marginal costs of producing an extra unit. If the marginal revenue exceeds the marginal cost, then the firm should produce the extra unit.

For example, at an output of 3 in [\[link\]](#), marginal revenue is 800 and marginal cost is 400, so producing this unit will clearly add to overall profits. At an output of 4, marginal revenue is 600 and marginal cost is 600, so producing this unit still means overall profits are unchanged. However, expanding output from 4 to 5 would involve a marginal revenue of 400 and a marginal cost of 700, so that fifth unit would actually reduce profits. Thus, the monopoly can tell from the marginal revenue and marginal cost that of the choices given in the table, the profit-maximizing level of output is 4.

Indeed, the monopoly could seek out the profit-maximizing level of output by increasing quantity by a small amount, calculating marginal revenue and marginal cost, and then either increasing output as long as marginal revenue exceeds marginal cost or reducing output if marginal cost exceeds marginal revenue. This process works without any need to calculate total revenue and total cost. Thus, a profit-maximizing monopoly should follow the rule of producing up to the quantity where marginal revenue is equal to marginal cost—that is,  $MR = MC$ .

### **Note:**

#### **Maximizing Profits**

If you find it counterintuitive that producing where marginal revenue equals marginal cost will maximize profits, working through the numbers will help.

Step 1. Remember that marginal cost is defined as the change in total cost from producing a small amount of additional output.

#### **Equation:**

$$MC = \frac{\text{change in total cost}}{\text{change in quantity produced}}$$

Step 2. Note that in [\[link\]](#), as output increases from 1 to 2 units, total cost increases from \$1500 to \$1800. As a result, the marginal cost of the second unit will be:

#### **Equation:**

$$\begin{aligned} MC &= \frac{\$1800 - \$1500}{1} \\ &= \$300 \end{aligned}$$

Step 3. Remember that, similarly, marginal revenue is the change in total revenue from selling a small amount of additional output.

#### **Equation:**

$$MR = \frac{\text{change in total revenue}}{\text{change in quantity sold}}$$

Step 4. Note that in [\[link\]](#), as output increases from 1 to 2 units, total revenue increases from \$1200 to \$2200. As a result, the marginal revenue of the second unit will be:

#### **Equation:**

$$\begin{aligned} MR &= \frac{\$2200 - \$1200}{1} \\ &= \$1000 \end{aligned}$$

Quantity	Marginal Revenue	Marginal Cost	Marginal Profit	Total Profit
1	1,200	1,500	-300	-300
2	1,000	300	700	400
3	800	400	400	800
4	600	600	0	800
5	400	700	-300	500
6	200	700	-500	0
7	0	1,400	-1,400	-1,400

### Marginal Revenue, Marginal Cost, Marginal and Total Profit

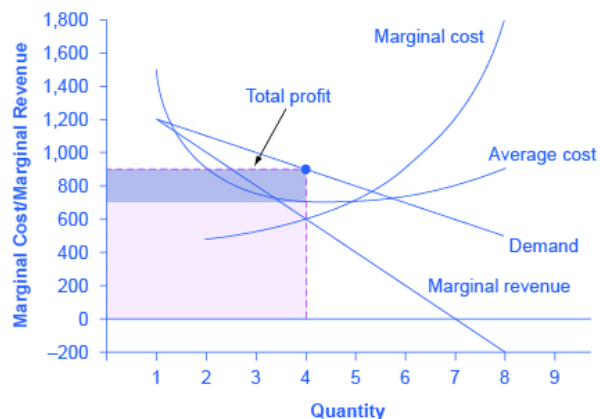
[\[link\]](#) repeats the marginal cost and marginal revenue data from [\[link\]](#), and adds two more columns: **Marginal profit** is the profitability of each additional unit sold. It is defined as marginal revenue minus marginal cost. Finally, total profit is the sum of marginal profits. As long as marginal profit is positive, producing more output will increase total profits. When marginal profit turns negative, producing more output will decrease total profits. Total profit is maximized where marginal revenue equals marginal cost. In this example, maximum profit occurs at 4 units of output.

A perfectly competitive firm will also find its profit-maximizing level of output where  $MR = MC$ . The key difference with a perfectly competitive firm is that in the case of perfect competition, marginal revenue is equal to price ( $MR = P$ ), while for a monopolist, marginal revenue is not equal to the price, because changes in quantity of output affect the price.

### Illustrating Monopoly Profits

It is straightforward to calculate profits of given numbers for total revenue and total cost. However, the size of monopoly profits can also be illustrated graphically with [\[link\]](#), which takes the marginal cost and marginal revenue curves from the previous exhibit and adds an average cost curve and the monopolist's perceived demand curve.

#### Illustrating Profits at the HealthPill Monopoly



This figure begins with the same marginal revenue and marginal cost curves from the HealthPill monopoly presented in [\[link\]](#). It

cost curves from the HealthPill monopoly presented in [\[link\]](#). It then adds an average cost curve and the demand curve faced by the monopolist. The HealthPill firm first chooses the quantity where  $MR = MC$ ; in this example, the quantity is 4. The monopolist then decides what price to charge by looking at the demand curve it faces. The large box, with quantity on the horizontal axis and marginal revenue on the vertical axis, shows total revenue for the firm. Total costs for the firm are shown by the lighter-shaded box, which is quantity on the horizontal axis and marginal cost of production on the vertical axis. The large total revenue box minus the smaller total cost box leaves the darkly shaded box that shows total profits. Since the price charged is above average cost, the firm is earning positive profits.

[\[link\]](#) illustrates the three-step process where a monopolist: selects the profit-maximizing quantity to produce; decides what price to charge; determines total revenue, total cost, and profit.

### **Step 1: The Monopolist Determines Its Profit-Maximizing Level of Output**

The firm can use the points on the demand curve  $D$  to calculate total revenue, and then, based on total revenue, calculate its marginal revenue curve. The profit-maximizing quantity will occur where  $MR = MC$ —or at the last possible point before marginal costs start exceeding marginal revenue. On [\[link\]](#),  $MR = MC$  occurs at an output of 4.

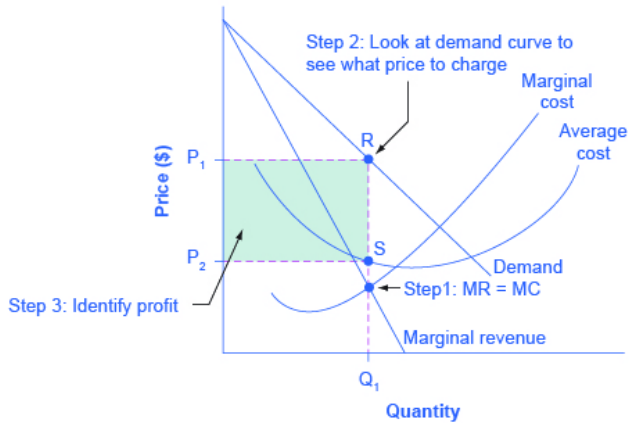
### **Step 2: The Monopolist Decides What Price to Charge**

The monopolist will charge what the market is willing to pay. A dotted line drawn straight up from the profit-maximizing quantity to the demand curve shows the profit-maximizing price. This price is above the average cost curve, which shows that the firm is earning profits.

### **Step 3: Calculate Total Revenue, Total Cost, and Profit**

Total revenue is the overall shaded box, where the width of the box is the quantity being sold and the height is the price. In [\[link\]](#), the bottom part of the shaded box, which is shaded more lightly, shows total costs; that is, quantity on the horizontal axis multiplied by average cost on the vertical axis. The larger box of total revenues minus the smaller box of total costs will equal profits, which is shown by the darkly shaded box. In a perfectly competitive market, the forces of entry would erode this profit in the long run. But a monopolist is protected by barriers to entry. In fact, one telltale sign of a possible monopoly is when a firm earns profits year after year, while doing more or less the same thing, without ever seeing those profits eroded by increased competition.

**How a Profit-Maximizing Monopoly Decides Price**



In Step 1, the monopoly chooses the profit-maximizing level of output  $Q_1$ , by choosing the quantity where  $MR = MC$ . In Step 2, the monopoly decides how much to charge for output level  $Q_1$  by drawing a line straight up from  $Q_1$  to point R on its perceived demand curve. Thus, the monopoly will charge a price ( $P_1$ ). In Step 3, the monopoly identifies its profit. Total revenue will be  $Q_1$  multiplied by  $P_1$ . Total cost will be  $Q_1$  multiplied by the average cost of producing  $Q_1$ , which is shown by point S on the average cost curve to be  $P_2$ . Profits will be the total revenue rectangle minus the total cost rectangle, shown by the shaded zone in the figure.

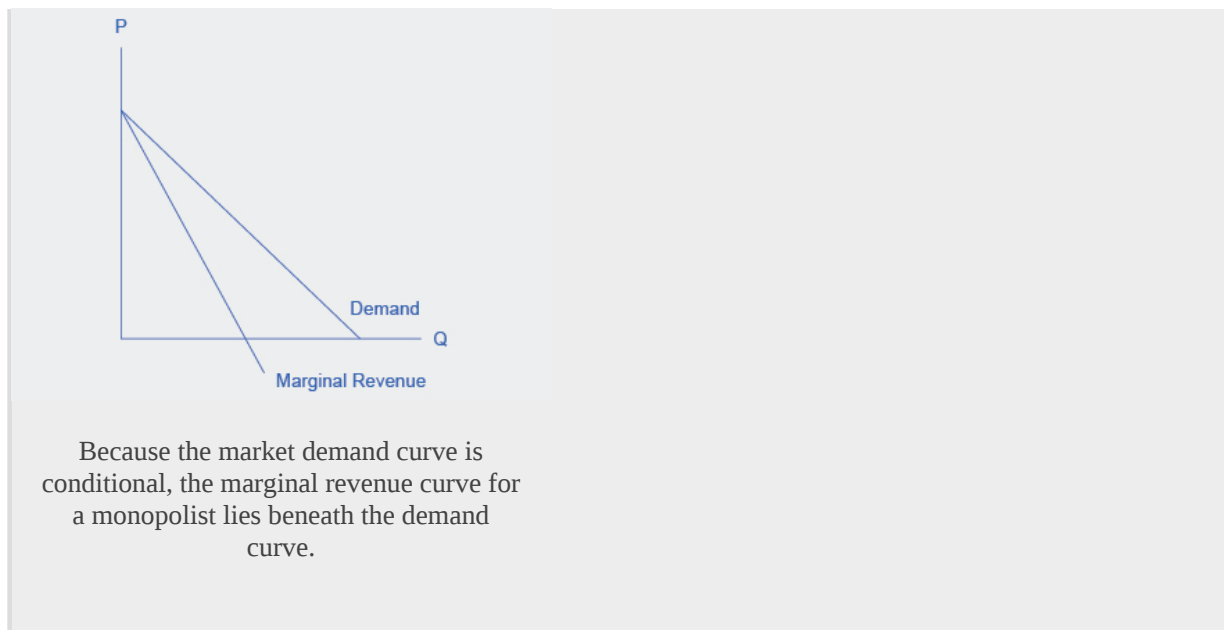
#### Note:

Why is a monopolist's marginal revenue always less than the price?

The marginal revenue curve for a monopolist always lies beneath the market demand curve. To understand why, think about increasing the quantity along the demand curve by one unit, so that you take one step down the demand curve to a slightly higher quantity but a slightly lower price. A demand curve is not sequential: It is not that first we sell  $Q_1$  at a higher price, and then we sell  $Q_2$  at a lower price. Rather, a demand curve is conditional: If we charge the higher price, we would sell  $Q_1$ . If, instead, we charge a lower price (on all the units that we sell), we would sell  $Q_2$ .

So when we think about increasing the quantity sold by one unit, marginal revenue is affected in two ways. First, we sell one additional unit at the new market price. Second, all the previous units, which could have been sold at the higher price, now sell for less. Because of the lower price on all units sold, the marginal revenue of selling a unit is less than the price of that unit—and the marginal revenue curve is below the demand curve. *Tip:* For a straight-line demand curve, MR and demand have the same vertical intercept. As output increases, marginal revenue decreases twice as fast as demand, so that the horizontal intercept of MR is halfway to the horizontal intercept of demand. You can see this in the [link](#).

The Monopolist's Marginal Revenue Curve versus Demand Curve



## The Inefficiency of Monopoly

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

**Allocative efficiency** is a social concept. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where  $P = MC$ , where the price ( $P$ ) is a measure of how much buyers value the good and the marginal cost ( $MC$ ) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If  $P > MC$ , then the marginal benefit to society (as measured by  $P$ ) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. But in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as can be seen by looking back at [\[link\]](#). Thus, consumers will suffer from a monopoly because a lower quantity will be sold in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: “The best of all monopoly profits is a quiet life.” He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When AT&T provided all of the local and long-distance phone service in the United States, along with manufacturing most of the phone equipment, the payment plans and types of phones did not change much. The old joke was that you could have any color phone you wanted, as long as it was black. But in 1982, AT&T was split up by government litigation into a number of local phone companies, a long-distance phone company,

and a phone equipment manufacturer. An explosion of innovation followed. Services like call waiting, caller ID, three-way calling, voice mail through the phone company, mobile phones, and wireless connections to the Internet all became available. A wide range of payment plans was offered, as well. It was no longer true that all phones were black; instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

**Note:****The Rest is History**

In the opening case, the East India Company and the Confederate States were presented as a monopoly or near monopoly provider of a good. Nearly every American schoolchild knows the result of the ‘unwelcome visit’ the ‘Mohawks’ bestowed upon Boston Harbor’s tea-bearing ships—the Boston Tea Party. Regarding the cotton industry, we also know Great Britain remained neutral during the Civil War, taking neither side during the conflict.

Did the monopoly nature of these businesses have unintended and historical consequences? Might the American Revolution have been deterred, if the East India Company had sailed the tea-bearing ships back to England? Might the southern states have made different decisions had they not been so confident “King Cotton” would force diplomatic recognition of the Confederate States of America? Of course, it is not possible to definitively answer these questions; after all we cannot roll back the clock and try a different scenario. We can, however, consider the monopoly nature of these businesses and the roles they played and hypothesize about what might have occurred under different circumstances.

Perhaps if there had been legal free tea trade, the colonists would have seen things differently; there was smuggled Dutch tea in the colonial market. If the colonists had been able to freely purchase Dutch tea, they would have paid lower prices and avoided the tax.

What about the cotton monopoly? With one in five jobs in Great Britain depending on Southern cotton and the Confederate States nearly the sole provider of that cotton, why did Great Britain remain neutral during the Civil War? At the beginning of the war, Britain simply drew down massive stores of cotton. These stockpiles lasted until near the end of 1862. Why did Britain not recognize the Confederacy at that point? Two reasons: The Emancipation Proclamation and new sources of cotton. Having outlawed slavery throughout the United Kingdom in 1833, it was politically impossible for Great Britain, empty cotton warehouses or not, to recognize, diplomatically, the Confederate States. In addition, during the two years it took to draw down the stockpiles, Britain expanded cotton imports from India, Egypt, and Brazil.

Monopoly sellers often see no threats to their superior marketplace position. In these examples did the power of the monopoly blind the decision makers to other possibilities? Perhaps. But, as they say, the rest is history.

**Key Concepts and Summary**

A monopolist is not a price taker, because when it decides what quantity to produce, it also determines the market price. For a monopolist, total revenue is relatively low at low quantities of output, because not much is being sold. Total revenue is also relatively low at very high quantities of output, because a very high quantity will sell only at a low price. Thus, total revenue for a monopolist will start low, rise, and then decline. The marginal revenue for a monopolist from selling additional units will decline. Each additional unit sold by a monopolist will push down the overall market price, and as more units are sold, this lower price applies to more and more units.

The monopolist will select the profit-maximizing level of output where  $MR = MC$ , and then charge the price for that quantity of output as determined by the market demand curve. If that price is above average cost, the monopolist earns positive profits.

Monopolists are not productively efficient, because they do not produce at the minimum of the average cost curve. Monopolists are not allocatively efficient, because they do not produce at the quantity where  $P = MC$ .

As a result, monopolists produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Monopolists also may lack incentives for innovation, because they need not fear entry.

## Self-Check Questions

### Exercise:

#### Problem:

Suppose demand for a monopoly's product falls so that its profit-maximizing price is below average variable cost. How much output should the firm supply? *Hint:* Draw the graph.

---

#### Solution:

If price falls below AVC, the firm will not be able to earn enough revenues even to cover its variable costs. In such a case, it will suffer a smaller loss if it shuts down and produces no output. By contrast, if it stayed in operation and produced the level of output where  $MR = MC$ , it would lose all of its fixed costs plus some variable costs. If it shuts down, it only loses its fixed costs.

### Exercise:

#### Problem:

Imagine a monopolist could charge a different price to every customer based on how much he or she were willing to pay. How would this affect monopoly profits?

---

#### Solution:

This scenario is called "perfect price discrimination." The result would be that the monopolist would produce more output, the same amount in fact as would be produced by a perfectly competitive industry. However, there would be no consumer surplus since each buyer is paying exactly what they think the product is worth. Therefore, the monopolist would be earning the maximum possible profits.

## Review Questions

### Exercise:

#### Problem:

How is the demand curve perceived by a perfectly competitive firm different from the demand curve perceived by a monopolist?

### Exercise:

#### Problem:

How does the demand curve perceived by a monopolist compare with the market demand curve?

### Exercise:

**Problem:** Is a monopolist a price taker? Explain briefly.

### Exercise:

**Problem:** What is the usual shape of a total revenue curve for a monopolist? Why?

### Exercise:

**Problem:** What is the usual shape of a marginal revenue curve for a monopolist? Why?

**Exercise:**

**Problem:**

How can a monopolist identify the profit-maximizing level of output if it knows its total revenue and total cost curves?

**Exercise:**

**Problem:**

How can a monopolist identify the profit-maximizing level of output if it knows its marginal revenue and marginal costs?

**Exercise:**

**Problem:**

When a monopolist identifies its profit-maximizing quantity of output, how does it decide what price to charge?

**Exercise:**

**Problem:** Is a monopolist allocatively efficient? Why or why not?

**Exercise:**

**Problem:**

How does the quantity produced and price charged by a monopolist compare to that of a perfectly competitive firm?

## Critical Thinking Questions

**Exercise:**

**Problem:**

Imagine that you are managing a small firm and thinking about entering the market of a monopolist. The monopolist is currently charging a high price, and you have calculated that you can make a nice profit charging 10% less than the monopolist. Before you go ahead and challenge the monopolist, what possibility should you consider for how the monopolist might react?

**Exercise:**

**Problem:**

If a monopoly firm is earning profits, how much would you expect these profits to be diminished by entry in the long run?

## Problems

**Exercise:**

**Problem:**

Draw the demand curve, marginal revenue, and marginal cost curves from [\[link\]](#), and identify the quantity of output the monopoly wishes to supply and the price it will charge. Suppose demand for the monopoly's product increases dramatically. Draw the new demand curve. What happens to the marginal revenue as a result of the increase in demand? What happens to the marginal cost curve? Identify the new profit-maximizing quantity and price. Does the answer make sense to you?

**Exercise:****Problem:**

Draw a monopolist's demand curve, marginal revenue, and marginal cost curves. Identify the monopolist's profit-maximizing output level. Now, think about a slightly higher level of output (say  $Q_0 + 1$ ). According to the graph, is there any consumer willing to pay more than the marginal cost of that new level of output? If so, what does this mean?

**References**

Aboukhadijeh, Feross. "Chapter 20: Girding for War - The North and the South, 1861-1865." *StudyNotes, Inc.* Accessed July 7, 2013. <http://www.apstudynotes.org/us-history/outlines/chapter-20-girding-for-war-the-north-and-the-south-1861-1865/>.

British Parliament. "(28 August 1833). Slavery Abolition Act 1833; Section LXIV." Accessed July 2013. [http://www.pdavis.nl/Legis\\_07.htm](http://www.pdavis.nl/Legis_07.htm).

Dattel, E. (nd). "Cotton and the Civil War." *Mississippi Historical Society*. Accessed July 2013. <http://mshistorynow.mdah.state.ms.us/articles/291/cotton-and-the-civil-war>.

Gartner. 2015. "Gartner Says Tablet Sales Continue to Be Slow in 2015." Accessed March 12, 2015. <http://www.gartner.com/newsroom/id/2954317>.

Grogan, David. 2015. "Federal Judge Finds AmEx's Anti-Steering Rule Violates Antitrust Law." American Booksellers Association. Accessed March 12, 2015. <http://www.bookweb.org/news/federal-judge-finds-amex%E2%80%99s-anti-steering-rule-violates-antitrust-law>.

Massachusetts Historical Society. "The Coming of the American Revolution 1764-1776: The Boston Tea Party." Retrieved from <http://www.masshist.org/revolution/teaparty.php>.

Massachusetts Historical Society. "Whereas our Nation." *The Massachusetts Gazette*, p. 2. Accessed July 2013 [http://www.masshist.org/revolution/image-viewer.php?old=1&item\\_id=457&img\\_step=1&nmask=1&mode=large](http://www.masshist.org/revolution/image-viewer.php?old=1&item_id=457&img_step=1&nmask=1&mode=large).

Pelegrin, William. 2015. "Judge Overrules Antitrust Case Against Google , Says Setting Default Search Engines Is Fair." Digital Trends. Accessed March 12, 2015. <http://www.digitaltrends.com/mobile/judge-tosses-out-google-antitrust-lawsuit/>.

**Glossary****allocative efficiency**

producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

**marginal profit**

profit of one more unit of output, computed as marginal revenue minus marginal cost

## Introduction to Monopolistic Competition and Oligopoly

class="introduction"

### Competing Brands?

The laundry  
detergent  
market is  
one that is  
characterize  
d neither as  
perfect  
competition  
nor  
monopoly.

(Credit:  
modification  
of work by  
Pixel  
Drip/Flickr  
Creative  
Commons)

**Note:****The Temptation to Defy the Law**

Laundry detergent and bags of ice—products of industries that seem pretty mundane, maybe even boring. Hardly! Both have been the center of clandestine meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

Around the same time, the top five Midwest ice makers (Home City Ice, Lang Ice, Tinley Ice, Sisler's Dairy, and Products of Ohio) had similar goals in mind when they secretly agreed to divide up the bagged ice market.

If both groups could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including the

European Union and the United States, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist? We will revisit this case later, to find out what happened.

**Note:**

**Introduction to Monopolistic Competition and Oligopoly**

In this chapter, you will learn about:

- Monopolistic Competition
- Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, price takers are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. Microsoft, for instance, has been considered a monopoly because of its domination of the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that could be described as **imperfectly competitive**? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly would. What will they do?

One type of imperfectly competitive market is called **monopolistic competition**. Monopolistically competitive markets feature a large number

of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of America in Minnesota, the largest shopping mall in the United States. In 2010, the Mall of America had 24 stores that sold women's "ready-to-wear" clothing (like Ann Taylor and Urban Outfitters), another 50 stores that sold clothing for both men and women (like Banana Republic, J. Crew, and Nordstrom's), plus 14 more stores that sold women's specialty clothing (like Motherhood Maternity and Victoria's Secret). Most of the markets that consumers encounter at the retail level are monopolistically competitive.

The other type of imperfectly competitive market is **oligopoly**. Oligopolistic markets are those dominated by a small number of firms. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the U.S. soft drink industry, which is dominated by Coca-Cola and Pepsi. Oligopolies are characterized by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this chapter, we first explore how monopolistically competitive firms will choose their profit-maximizing level of output. We will then discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

## Monopolistic Competition

By the end of this section, you will be able to:

- Explain the significance of differentiated products
- Describe how a monopolistic competitor chooses price and quantity
- Discuss entry, exit, and efficiency as they pertain to monopolistic competition
- Analyze how advertising can impact monopolistic competition

Monopolistic competition involves many firms competing against each other, but selling products that are distinctive in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell different kinds of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. There are over 600,000 restaurants in the United States. When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term “monopolistic competition” captures this mixture of mini-monopoly and tough competition, and the following Clear It Up feature introduces its derivation.

### Note:

Who invented the theory of imperfect competition?

The theory of imperfect competition was developed by two economists independently but simultaneously in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics where she became a prominent Keynesian, and later a post-Keynesian economist. (See the [Welcome to Economics!](#) and [The Keynesian Perspective](#) chapters for more on Keynes.)

## Differentiated Products

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which the product is sold, intangible aspects of the product, and perceptions of the product. Products that are distinctive in one of these ways are called **differentiated products**.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle, nonstick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. The location of a firm can also create a difference between producers. For example, a gas station located at a heavily traveled intersection can probably sell more gas, because more cars drive by that corner. A supplier to an automobile manufacturer may find that it is an advantage to locate close to the car factory.

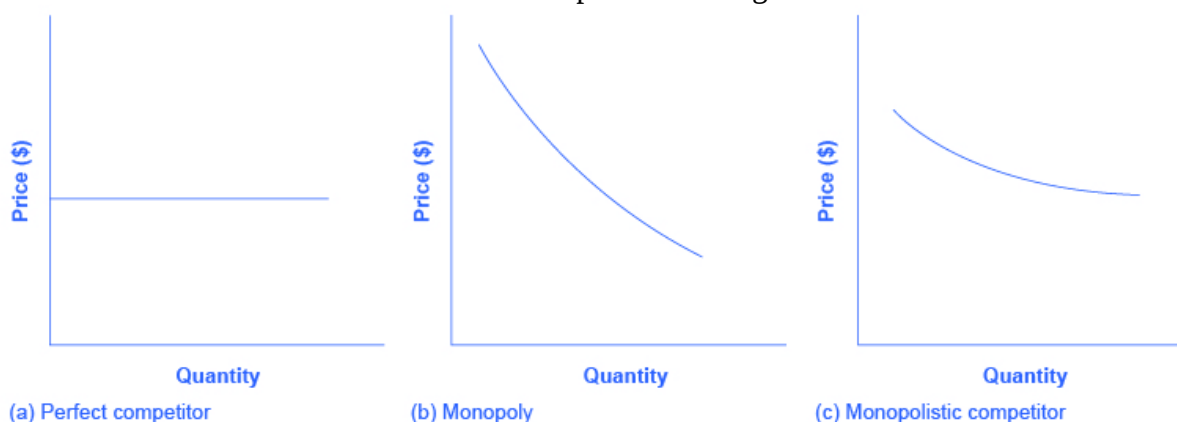
Intangible aspects can differentiate a product, too. Some intangible aspects may be promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, product differentiation may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of beer or cigarettes if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these intangible preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

## Perceived Demand for a Monopolistic Competitor

A monopolistically competitive firm perceives a demand for its goods that is an intermediate case between monopoly and competition. [\[link\]](#) offers a reminder that the demand curve as faced by a perfectly competitive firm is perfectly elastic or flat, because the perfectly competitive firm can sell any quantity it wishes at the prevailing market price. In contrast, the demand curve, as faced by a monopolist, is the market demand curve, since a monopolist is the only firm in the market, and hence is downward sloping.

Perceived Demand for Firms in Different Competitive Settings



The demand curve faced by a perfectly competitive firm is perfectly elastic, meaning it can sell all the output it wishes at the prevailing market price. The demand curve faced by a monopoly is the market demand. It can sell more output only by decreasing the price it charges. The demand curve faced by a monopolistically competitive firm falls in between.

The demand curve as faced by a monopolistic competitor is not flat, but rather downward-sloping, which means that the monopolistic competitor can raise its price without losing all of its customers or lower the price and gain more customers. Since there are substitutes, the demand curve facing a monopolistically competitive firm is more elastic than that of a monopoly where

there are no close substitutes. If a monopolist raises its price, some consumers will choose not to purchase its product—but they will then need to buy a completely different product. However, when a monopolistic competitor raises its price, some consumers will choose not to purchase the product at all, but others will choose to buy a similar product from another firm. If a monopolistic competitor raises its price, it will not lose as many customers as would a perfectly competitive firm, but it will lose more customers than would a monopoly that raised its prices.

At a glance, the demand curves faced by a monopoly and by a monopolistic competitor look similar—that is, they both slope down. But the underlying economic meaning of these perceived demand curves is different, because a monopolist faces the market demand curve and a monopolistic competitor does not. Rather, a monopolistically competitive firm's demand curve is but one of many firms that make up the “before” market demand curve. Are you following? If so, how would you categorize the market for golf balls? Take a swing, then see the following Clear It Up feature.

**Note:**

**Are golf balls really differentiated products?**

Monopolistic competition refers to an industry that has more than a few firms, each offering a product which, from the consumer's perspective, is different from its competitors. The U.S. Golf Association runs a laboratory that tests 20,000 golf balls a year. There are strict rules for what makes a golf ball legal. The weight of a golf ball cannot exceed 1.620 ounces and its diameter cannot be less than 1.680 inches (which is a weight of 45.93 grams and a diameter of 42.67 millimeters, in case you were wondering). The balls are also tested by being hit at different speeds. For example, the distance test involves having a mechanical golfer hit the ball with a titanium driver and a swing speed of 120 miles per hour. As the testing center explains: “The USGA system then uses an array of sensors that accurately measure the flight of a golf ball during a short, indoor trajectory from a ball launcher. From this flight data, a computer calculates the lift and drag forces that are generated by the speed, spin, and dimple pattern of the ball. ... The distance limit is 317 yards.”

Over 1800 golf balls made by more than 100 companies meet the USGA standards. The balls do differ in various ways, like the pattern of dimples on the ball, the types of plastic used on the cover and in the cores, and so on. Since all balls need to conform to the USGA tests, they are much more alike than different. In other words, golf ball manufacturers are monopolistically competitive.

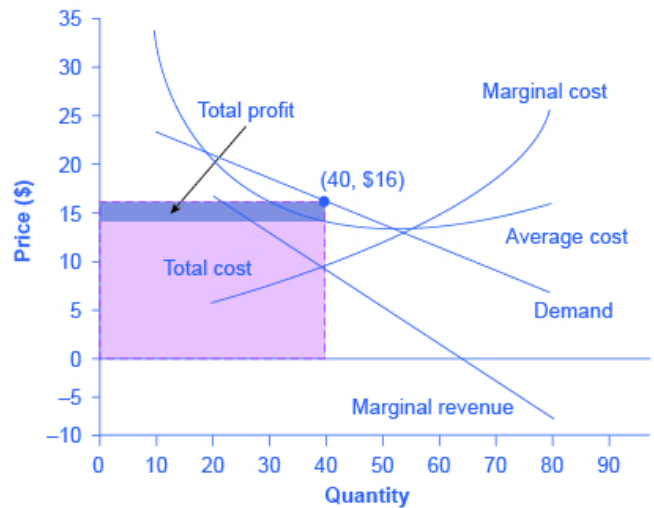
However, retail sales of golf balls are about \$500 million per year, which means that a lot of large companies have a powerful incentive to persuade players that golf balls are highly differentiated and that it makes a huge difference which one you choose. Sure, Tiger Woods can tell the difference. For the average duffer (golf-speak for a “mediocre player”) who plays a few times a summer—and who loses a lot of golf balls to the woods and lake and needs to buy new ones—most golf balls are pretty much indistinguishable.

## **How a Monopolistic Competitor Chooses Price and Quantity**

The monopolistically competitive firm decides on its profit-maximizing quantity and price in much the same way as a monopolist. A monopolistic competitor, like a monopolist, faces a downward-sloping demand curve, and so it will choose some combination of price and quantity along its perceived demand curve.

As an example of a profit-maximizing monopolistic competitor, consider the Authentic Chinese Pizza store, which serves pizza with cheese, sweet and sour sauce, and your choice of vegetables and meats. Although Authentic Chinese Pizza must compete against other pizza businesses and restaurants, it has a differentiated product. The firm’s perceived demand curve is downward sloping, as shown in [\[link\]](#) and the first two columns of [\[link\]](#).

How a Monopolistic Competitor Chooses its Profit Maximizing Output and Price



To maximize profits, the Authentic Chinese Pizza shop would choose a quantity where marginal revenue equals marginal cost, or  $Q$  where  $MR = MC$ . Here it would choose a quantity of 40 and a price of \$16.

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
10	\$23	\$230	-	\$340	-	\$34
20	\$20	\$400	\$17	\$400	\$6	\$20
30	\$18	\$540	\$14	\$480	\$8	\$16

Quantity	Price	Total Revenue	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
40	\$16	\$640	\$10	\$580	\$10	\$14.50
50	\$14	\$700	\$6	\$700	\$12	\$14
60	\$12	\$720	\$2	\$840	\$14	\$14
70	\$10	\$700	-\$2	\$1,020	\$18	\$14.57
80	\$8	\$640	-\$6	\$1,280	\$26	\$16

### Revenue and Cost Schedule

The combinations of price and quantity at each point on the demand curve can be multiplied to calculate the total revenue that the firm would receive, which is shown in the third column of [\[link\]](#). The fourth column, marginal revenue, is calculated as the change in total revenue divided by the change in quantity. The final columns of [\[link\]](#) show total cost, marginal cost, and average cost. As always, marginal cost is calculated by dividing the change in total cost by the change in quantity, while average cost is calculated by dividing total cost by quantity. The following Work It Out feature shows how these firms calculate how much of its product to supply at what price.

#### Note:

##### How a Monopolistic Competitor Determines How Much to Produce and at What Price

The process by which a monopolistic competitor chooses its profit-maximizing quantity and price resembles closely how a monopoly makes these decisions process. First, the firm selects the profit-maximizing quantity to produce. Then the firm decides what price to charge for that quantity.

Step 1. The monopolistic competitor determines its profit-maximizing level of output. In this case, the Authentic Chinese Pizza company will determine the profit-maximizing quantity to produce by considering its marginal revenues and marginal costs. Two scenarios are possible:

- If the firm is producing at a quantity of output where marginal revenue exceeds marginal cost, then the firm should keep expanding production, because each marginal unit is adding to profit by bringing in more revenue than its cost. In this way, the firm will produce up to the quantity where  $MR = MC$ .
- If the firm is producing at a quantity where marginal costs exceed marginal revenue, then each marginal unit is costing more than the revenue it brings in, and the firm will increase its profits by reducing the quantity of output until  $MR = MC$ .

In this example, MR and MC intersect at a quantity of 40, which is the profit-maximizing level of output for the firm.

Step 2. The monopolistic competitor decides what price to charge. When the firm has determined its profit-maximizing quantity of output, it can then look to its perceived demand

curve to find out what it can charge for that quantity of output. On the graph, this process can be shown as a vertical line reaching up through the profit-maximizing quantity until it hits the firm's perceived demand curve. For Authentic Chinese Pizza, it should charge a price of \$16 per pizza for a quantity of 40.

Once the firm has chosen price and quantity, it's in a position to calculate total revenue, total cost, and profit. At a quantity of 40, the price of \$16 lies above the average cost curve, so the firm is making economic profits. From [\[link\]](#) we can see that, at an output of 40, the firm's total revenue is \$640 and its total cost is \$580, so profits are \$60. In [\[link\]](#), the firm's total revenues are the rectangle with the quantity of 40 on the horizontal axis and the price of \$16 on the vertical axis. The firm's total costs are the light shaded rectangle with the same quantity of 40 on the horizontal axis but the average cost of \$14.50 on the vertical axis. Profits are total revenues minus total costs, which is the shaded area above the average cost curve.

Although the process by which a monopolistic competitor makes decisions about quantity and price is similar to the way in which a monopolist makes such decisions, two differences are worth remembering. First, although both a monopolist and a monopolistic competitor face downward-sloping demand curves, the monopolist's perceived demand curve is the market demand curve, while the perceived demand curve for a monopolistic competitor is based on the extent of its product differentiation and how many competitors it faces. Second, a monopolist is surrounded by barriers to entry and need not fear entry, but a monopolistic competitor who earns profits must expect the entry of firms with similar, but differentiated, products.

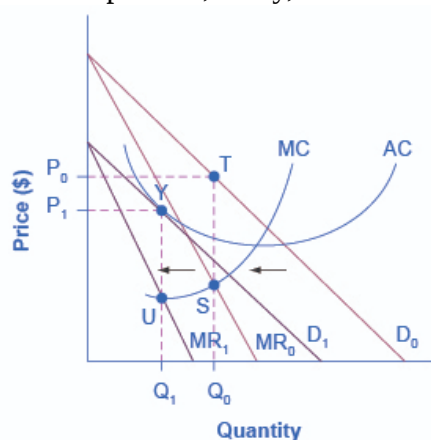
## Monopolistic Competitors and Entry

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A gas station with a great location must worry that other gas stations might open across the street or down the road—and perhaps the new gas stations will sell coffee or have a carwash or some other attraction to lure customers. A successful restaurant with a unique barbecue sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A laundry detergent with a great reputation for quality must be concerned that other competitors may seek to build their own reputations.

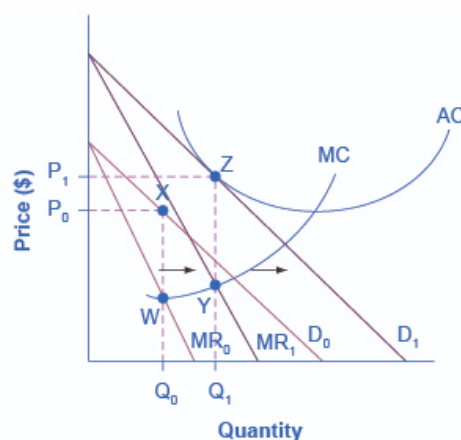
The entry of other firms into the same general market (like gas, restaurants, or detergent) shifts the demand curve faced by a monopolistically competitive firm. As more firms enter the market, the quantity demanded at a given price for any particular firm will decline, and the firm's perceived demand curve will shift to the left. As a firm's perceived demand curve shifts to the left, its marginal revenue curve will shift to the left, too. The shift in marginal revenue will change the profit-maximizing quantity that the firm chooses to produce, since marginal revenue will then equal marginal cost at a lower quantity.

[\[link\]](#) (a) shows a situation in which a monopolistic competitor was earning a profit with its original perceived demand curve ( $D_0$ ). The intersection of the marginal revenue curve ( $MR_0$ ) and marginal cost curve ( $MC$ ) occurs at point S, corresponding to quantity  $Q_0$ , which is associated on the demand curve at point T with price  $P_0$ . The combination of price  $P_0$  and quantity  $Q_0$  lies above the average cost curve, which shows that the firm is earning positive economic profits.

## Monopolistic Competition, Entry, and Exit



(a) Profit induces entry; shift to zero profit



(b) Loss induces exit; shift to zero profit

(a) At  $P_0$  and  $Q_0$ , the monopolistically competitive firm shown in this figure is making a positive economic profit. This is clear because if you follow the dotted line above  $Q_0$ , you can see that price is above average cost. Positive economic profits attract competing firms to the industry, driving the original firm's demand down to  $D_1$ . At the new equilibrium quantity ( $P_1, Q_1$ ), the original firm is earning zero economic profits, and entry into the industry ceases. In (b) the opposite occurs. At  $P_0$  and  $Q_0$ , the firm is losing money. If you follow the dotted line above  $Q_0$ , you can see that average cost is above price. Losses induce firms to leave the industry. When they do, demand for the original firm rises to  $D_1$ , where once again the firm is earning zero economic profit.

Unlike a monopoly, with its high barriers to entry, a monopolistically competitive firm with positive economic profits will attract competition. When another competitor enters the market, the original firm's perceived demand curve shifts to the left, from  $D_0$  to  $D_1$ , and the associated marginal revenue curve shifts from  $MR_0$  to  $MR_1$ . The new profit-maximizing output is  $Q_1$ , because the intersection of the  $MR_1$  and  $MC$  now occurs at point U. Moving vertically up from that quantity on the new demand curve, the optimal price is at  $P_1$ .

As long as the firm is earning positive economic profits, new competitors will continue to enter the market, reducing the original firm's demand and marginal revenue curves. The long-run equilibrium is shown in the figure at point Y, where the firm's perceived demand curve touches the average cost curve. When price is equal to average cost, economic profits are zero. Thus, although a monopolistically competitive firm may earn positive economic profits in the short term, the process of new entry will drive down economic profits to zero in the long run. Remember that zero economic profit is not equivalent to zero accounting profit. A zero economic profit means the firm's accounting profit is equal to what its resources could earn in their next best use. [\[link\]](#) (b) shows the reverse situation, where a monopolistically competitive firm is originally losing money. The adjustment to long-run equilibrium is analogous to the previous example. The economic losses lead to firms exiting, which will result in increased demand for this particular firm, and consequently lower losses. Firms exit up to the point where

there are no more losses in this market, for example when the demand curve touches the average cost curve, as in point Z.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

## Monopolistic Competition and Efficiency

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays productive efficiency: goods are being produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short run and in the long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set  $MR = MC$ —and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When  $P > MC$ , which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case to a lesser degree. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry. See the following Clear It Up feature for more detail on the impact of demand shifts.

### **Note:**

**Why does a shift in perceived demand cause a shift in marginal revenue?**

The combinations of price and quantity at each point on a firm's perceived demand curve are used to calculate total revenue for each combination of price and quantity. This information on total revenue is then used to calculate marginal revenue, which is the change in total revenue divided by the change in quantity. A change in perceived demand will change total revenue at every quantity of output and in turn, the change in total revenue will shift marginal revenue at each quantity of output. Thus, when entry occurs in a monopolistically competitive industry, the perceived demand curve for each firm will shift to the left, because a smaller quantity will be demanded at any given price. Another way of interpreting this shift in demand is to notice that, for each quantity sold, a lower price will be charged. Consequently, the marginal revenue will

be lower for each quantity sold—and the marginal revenue curve will shift to the left as well. Conversely, exit causes the perceived demand curve for a monopolistically competitive firm to shift to the right and the corresponding marginal revenue curve to shift right, too.

A monopolistically competitive industry does not display productive and allocative efficiency in either the short run, when firms are making economic profits and losses, nor in the long run, when firms are earning zero profits.

## **The Benefits of Variety and Product Differentiation**

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Many people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Many people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as happy with a smaller range of differentiated products produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers. Read the following Clear It Up feature for a discussion on the role that advertising plays in monopolistic competition.

### **Note:**

**How does advertising impact monopolistic competition?**

The U.S. economy spent about \$180.12 billion on advertising in 2014, according to eMarketer.com. Roughly one third of this was television advertising, and another third was divided roughly equally between Internet, newspapers, and radio. The remaining third was divided up between direct mail, magazines, telephone directory yellow pages, and billboards. Mobile devices are increasing the opportunities for advertisers.

Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from the products of another firm. In the framework of monopolistic

competition, there are two ways to conceive of how advertising works: either advertising causes a firm's perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm's product to increase (that is, it causes the firm's perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

"It may happen that expenditures on advertisement made by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from the other, the total result is the same as it would have been if neither had made any effort at all."

## Key Concepts and Summary

Monopolistic competition refers to a market where many firms sell differentiated products. Differentiated products can arise from characteristics of the good or service, location from which the product is sold, intangible aspects of the product, and perceptions of the product.

The perceived demand curve for a monopolistically competitive firm is downward-sloping, which shows that it is a price maker and chooses a combination of price and quantity. However, the perceived demand curve for a monopolistic competitor is more elastic than the perceived demand curve for a monopolist, because the monopolistic competitor has direct competition, unlike the pure monopolist. A profit-maximizing monopolistic competitor will seek out the quantity where marginal revenue is equal to marginal cost. The monopolistic competitor will produce that level of output and charge the price that is indicated by the firm's demand curve.

If the firms in a monopolistically competitive industry are earning economic profits, the industry will attract entry until profits are driven down to zero in the long run. If the firms in a monopolistically competitive industry are suffering economic losses, then the industry will experience exit of firms until economic profits are driven up to zero in the long run.

A monopolistically competitive firm is not productively efficient because it does not produce at the minimum of its average cost curve. A monopolistically competitive firm is not allocatively efficient because it does not produce where  $P = MC$ , but instead produces where  $P > MC$ . Thus, a monopolistically competitive firm will tend to produce a lower quantity at a higher cost and to charge a higher price than a perfectly competitive firm.

Monopolistically competitive industries do offer benefits to consumers in the form of greater variety and incentives for improved products and services. There is some controversy over whether a market-oriented economy generates too much variety.

## Self-Check Questions

**Exercise:****Problem:**

Suppose that, due to a successful advertising campaign, a monopolistic competitor experiences an increase in demand for its product. How will that affect the price it charges and the quantity it supplies?

---

**Solution:**

An increase in demand will manifest itself as a rightward shift in the demand curve, and a rightward shift in marginal revenue. The shift in marginal revenue will cause a movement up the marginal cost curve to the new intersection between MR and MC at a higher level of output. The new price can be read by drawing a line up from the new output level to the new demand curve, and then over to the vertical axis. The new price should be higher. The increase in quantity will cause a movement along the average cost curve to a possibly higher level of average cost. The price, though, will increase more, causing an increase in total profits.

**Exercise:****Problem:**

Continuing with the scenario outlined in question 1, in the long run, the positive economic profits earned by the monopolistic competitor will attract a response either from existing firms in the industry or firms outside. As those firms capture the original firm's profit, what will happen to the original firm's profit-maximizing price and output levels?

---

**Solution:**

As long as the original firm is earning positive economic profits, other firms will respond in ways that take away the original firm's profits. This will manifest itself as a decrease in demand for the original firm's product, a decrease in the firm's profit-maximizing price and a decrease in the firm's profit-maximizing level of output, essentially unwinding the process described in the answer to question 1. In the long-run equilibrium, all firms in monopolistically competitive markets will earn zero economic profits.

**Review Questions****Exercise:****Problem:**

What is the relationship between product differentiation and monopolistic competition?

**Exercise:****Problem:**

How is the perceived demand curve for a monopolistically competitive firm different from the perceived demand curve for a monopoly or a perfectly competitive firm?

**Exercise:**

**Problem:**

How does a monopolistic competitor choose its profit-maximizing quantity of output and price?

**Exercise:**

**Problem:**

How can a monopolistic competitor tell whether the price it is charging will cause the firm to earn profits or experience losses?

**Exercise:**

**Problem:**

If the firms in a monopolistically competitive market are earning economic profits or losses in the short run, would you expect them to continue doing so in the long run? Why?

**Exercise:**

**Problem:**

Is a monopolistically competitive firm productively efficient? Is it allocatively efficient? Why or why not?

## **Critical Thinking Questions**

**Exercise:**

**Problem:**

Aside from advertising, how can monopolistically competitive firms increase demand for their products?

**Exercise:**

**Problem:**

Make a case for why monopolistically competitive industries never reach long-run equilibrium.

**Exercise:**

**Problem:**

Would you rather have efficiency or variety? That is, one opportunity cost of the variety of products we have is that each product costs more per unit than if there were only one kind of product of a given type, like shoes. Perhaps a better question is, "What is the right amount of variety? Can there be too many varieties of shoes, for example?"

## Problems

### Exercise:

#### Problem:

Andrea's Day Spa began to offer a relaxing aromatherapy treatment. The firm asks you how much to charge to maximize profits. The demand curve for the treatments is given by the first two columns in [\[link\]](#); its total costs are given in the third column. For each level of output, calculate total revenue, marginal revenue, average cost, and marginal cost. What is the profit-maximizing level of output for the treatments and how much will the firm earn in profits?

Price	Quantity	TC
\$25.00	0	\$130
\$24.00	10	\$275
\$23.00	20	\$435
\$22.50	30	\$610
\$22.00	40	\$800
\$21.60	50	\$1,005
\$21.20	60	\$1,225

## References

Kantar Media. "Our Insights: Infographic—U.S. Advertising Year End Trends Report 2012." Accessed October 17, 2013. <http://kantarmedia.us/insight-center/reports/infographic-us-advertising-year-end-trends-report-2012>.

Statistica.com. 2015. "Number of Restaurants in the United States from 2011 to 2014." Accessed March 27, 2015. <http://www.statista.com/statistics/244616/number-of-qsr-fsr-chain-independent-restaurants-in-the-us/>.

## Glossary

differentiated product

a product that is perceived by consumers as distinctive in some way

imperfectly competitive

firms and organizations that fall between the extremes of monopoly and perfect competition

monopolistic competition

many firms competing to sell similar but differentiated products

oligopoly

when a few large firms have all or most of the sales in an industry

## Oligopoly

By the end of this section, you will be able to:

- Explain why and how oligopolies exist
- Contrast collusion and competition
- Interpret and analyze the prisoner's dilemma diagram
- Evaluate the tradeoffs of imperfect competition

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. Oligopoly arises when a small number of large firms have all or most of the sales in an industry. Examples of oligopoly abound and include the auto industry, cable television, and commercial air travel. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists collude with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. Oligopolies are typically characterized by mutual interdependence where various decisions such as output, price, advertising, and so on, depend on the decisions of the other firm(s). Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

### **Why Do Oligopolies Exist?**

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a natural monopoly will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates the barrier to entry, which led to the Boeing-Airbus oligopoly for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task.

## **Collusion or Competition?**

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry output, charge a higher price, and divide up the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called **collusion**. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a **cartel**. See the following Clear It Up feature for a more in-depth analysis of the difference between the two.

**Note:**

**Collusion versus cartels: How can I tell which is which?**

In the United States, as well as many other countries, it is illegal for firms to collude since collusion is anti-competitive behavior, which is a violation of antitrust law. Both the Antitrust Division of the Justice Department and the Federal Trade Commission have responsibilities for preventing collusion in the United States.

The problem of enforcement is finding hard evidence of collusion. Cartels are formal agreements to collude. Because cartel agreements provide evidence of collusion, they are rare in the United States. Instead, most collusion is tacit, where firms implicitly reach an understanding that competition is bad for profits.

The desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits has been well understood by economists. Adam Smith wrote in *Wealth of Nations* in 1776: “People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.”

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. Indeed, a small handful of oligopoly firms may end up competing so fiercely that they all end up earning zero economic profits—as if they were perfect competitors.

## **The Prisoner’s Dilemma**

Because of the complexity of oligopoly, which is the result of mutual interdependence among firms, there is no single, generally-accepted theory of how oligopolies behave, in the same way that we have theories for all the other market structures. Instead, economists use **game theory**, a branch of

mathematics that analyzes situations in which players must make decisions and then receive payoffs based on what other players decide to do. Game theory has found widespread applications in the social sciences, as well as in business, law, and military strategy.

The **prisoner's dilemma** is a scenario in which the gains from cooperation are larger than the rewards from pursuing self-interest. It applies well to oligopoly. The story behind the prisoner's dilemma goes like this:

"Two co-conspiratorial criminals are arrested. When they are taken to the police station, they refuse to say anything and are put in separate interrogation rooms. Eventually, a police officer enters the room where Prisoner A is being held and says: "You know what? Your partner in the other room is confessing. So your partner is going to get a light prison sentence of just one year, and because you're remaining silent, the judge is going to stick you with eight years in prison. Why don't you get smart? If you confess, too, we'll cut your jail time down to five years, and your partner will get five years, also." Over in the next room, another police officer is giving exactly the same speech to Prisoner B. What the police officers do not say is that if both prisoners remain silent, the evidence against them is not especially strong, and the prisoners will end up with only two years in jail each."

The game theory situation facing the two prisoners is shown in [\[link\]](#). To understand the dilemma, first consider the choices from Prisoner A's point of view. If A believes that B will confess, then A ought to confess, too, so as to not get stuck with the eight years in prison. But if A believes that B will not confess, then A will be tempted to act selfishly and confess, so as to serve only one year. The key point is that A has an incentive to confess regardless of what choice B makes! B faces the same set of choices, and thus will have an incentive to confess regardless of what choice A makes. Confess is considered the dominant strategy or the strategy an individual (or firm) will pursue regardless of the other individual's (or firm's) decision. The result is that if prisoners pursue their own self-interest, both are likely to confess, and end up doing a total of 10 years of jail time between them.

		Prisoner B	
		Remain Silent (cooperate with other prisoner)	Confess (do not cooperate with other prisoner)
Prisoner A	Remain Silent (cooperate with other prisoner)	A gets 2 years, B gets 2 years	A gets 8 years, B gets 1 year
	Confess (do not cooperate with other prisoner)	A gets 1 year, B gets 8 years	A gets 5 years B gets 5 years

### The Prisoner's Dilemma Problem

The game is called a dilemma because if the two prisoners had cooperated by both remaining silent, they would only have had to serve a total of four years of jail time between them. If the two prisoners can work out some way of cooperating so that neither one will confess, they will both be better off than if they each follow their own individual self-interest, which in this case leads straight into longer jail terms.

### The Oligopoly Version of the Prisoner's Dilemma

The members of an oligopoly can face a prisoner's dilemma, also. If each of the oligopolists cooperates in holding down output, then high monopoly profits are possible. Each oligopolist, however, must worry that while it is holding down output, other firms are taking advantage of the high price by raising output and earning higher profits. [\[link\]](#) shows the prisoner's dilemma for a two-firm oligopoly—known as a **duopoly**. If Firms A and B both agree to hold down output, they are acting together as a monopoly and

will each earn \$1,000 in profits. However, both firms' dominant strategy is to increase output, in which case each will earn \$400 in profits.

		<b>Firm B</b>	
		Hold Down Output (cooperate with other firm)	Increase Output (do not cooperate with other firm)
<b>Firm A</b>	Hold Down Output (cooperate with other firm)	A gets \$1,000, B gets \$1,000	A gets \$200, B gets \$1,500
	Increase Output (do not cooperate with other firm)	A gets \$1,500, B gets \$200	A gets \$400, B gets \$400

### A Prisoner's Dilemma for Oligopolists

Can the two firms trust each other? Consider the situation of Firm A:

- If A thinks that B will cheat on their agreement and increase output, then A will increase output, too, because for A the profit of \$400 when both firms increase output (the bottom right-hand choice in [\[link\]](#)) is better than a profit of only \$200 if A keeps output low and B raises output (the upper right-hand choice in the table).
- If A thinks that B will cooperate by holding down output, then A may seize the opportunity to earn higher profits by raising output. After all, if B is going to hold down output, then A can earn \$1,500 in profits by

expanding output (the bottom left-hand choice in the table) compared with only \$1,000 by holding down output as well (the upper left-hand choice in the table).

Thus, firm A will reason that it makes sense to expand output if B holds down output and that it also makes sense to expand output if B raises output. Again, B faces a parallel set of decisions.

The result of this prisoner's dilemma is often that even though A and B could make the highest combined profits by cooperating in producing a lower level of output and acting like a monopolist, the two firms may well end up in a situation where they each increase output and earn only \$400 each in profits. The following Clear It Up feature discusses one cartel scandal in particular.

**Note:**

What is the Lysine cartel?

Lysine, a \$600 million-a-year industry, is an amino acid used by farmers as a feed additive to ensure the proper growth of swine and poultry. The primary U.S. producer of lysine is Archer Daniels Midland (ADM), but several other large European and Japanese firms are also in this market. For a time in the first half of the 1990s, the world's major lysine producers met together in hotel conference rooms and decided exactly how much each firm would sell and what it would charge. The U.S. Federal Bureau of Investigation (FBI), however, had learned of the cartel and placed wire taps on a number of their phone calls and meetings.

From FBI surveillance tapes, following is a comment that Terry Wilson, president of the corn processing division at ADM, made to the other lysine producers at a 1994 meeting in Mona, Hawaii:

"I wanna go back and I wanna say something very simple. If we're going to trust each other, okay, and if I'm assured that I'm gonna get 67,000 tons by the year's end, we're gonna sell it at the prices we agreed to . . . The only thing we need to talk about there because we are gonna get manipulated by these [expletive] buyers—they can be smarter than us if we let them be smarter. . . . They [the customers] are not your friend. They are not my friend. And we gotta have 'em, but they are not my friends. You are

my friend. I wanna be closer to you than I am to any customer. Cause you can make us ... money. ... And all I wanna tell you again is let's—let's put the prices on the board. Let's all agree that's what we're gonna do and then walk out of here and do it. "

The price of lysine doubled while the cartel was in effect. Confronted by the FBI tapes, Archer Daniels Midland pled guilty in 1996 and paid a fine of \$100 million. A number of top executives, both at ADM and other firms, later paid fines of up to \$350,000 and were sentenced to 24–30 months in prison.

In another one of the FBI recordings, the president of Archer Daniels Midland told an executive from another competing firm that ADM had a slogan that, in his words, had “penetrated the whole company.” The company president stated the slogan this way: “Our competitors are our friends. Our customers are the enemy.” That slogan could stand as the motto of cartels everywhere.

## **How to Enforce Cooperation**

How can parties who find themselves in a prisoner's dilemma situation avoid the undesired outcome and cooperate with each other? The way out of a prisoner's dilemma is to find a way to penalize those who do not cooperate.

Perhaps the easiest approach for colluding oligopolists, as you might imagine, would be to sign a contract with each other that they will hold output low and keep prices high. If a group of U.S. companies signed such a contract, however, it would be illegal. Certain international organizations, like the nations that are members of the Organization of Petroleum Exporting Countries (OPEC), have signed international agreements to act like a monopoly, hold down output, and keep prices high so that all of the countries can make high profits from oil exports. Such agreements, however, because they fall in a gray area of international law, are not legally enforceable. If Nigeria, for example, decides to start cutting prices and selling more oil, Saudi Arabia cannot sue Nigeria in court and force it to stop.

**Note:**

Visit the Organization of the Petroleum Exporting Countries [website](#) and learn more about its history and how it defines itself.



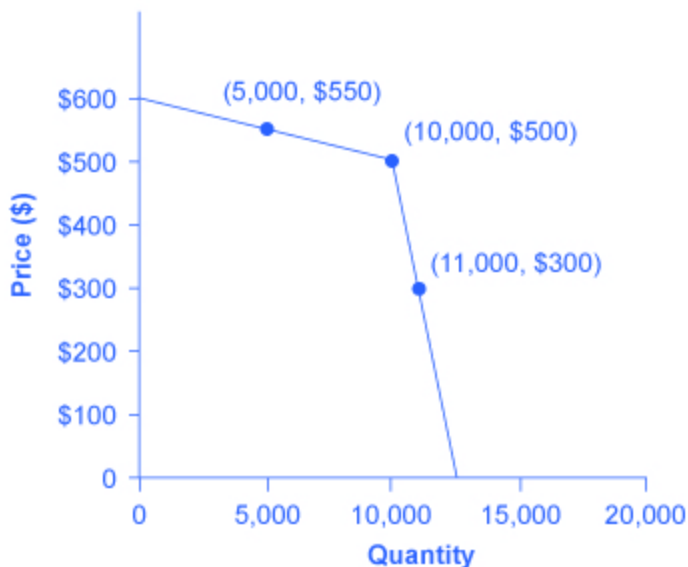
Because oligopolists cannot sign a legally enforceable contract to act like a monopoly, the firms may instead keep close tabs on what other firms are producing and charging. Alternatively, oligopolists may choose to act in a way that generates pressure on each firm to stick to its agreed quantity of output.

One example of the pressure these firms can exert on one another is the **kinked demand curve**, in which competing oligopoly firms commit to match price cuts, but not price increases. This situation is shown in [\[link\]](#). Say that an oligopoly airline has agreed with the rest of a cartel to provide a quantity of 10,000 seats on the New York to Los Angeles route, at a price of \$500. This choice defines the kink in the firm's perceived demand curve. The reason that the firm faces a kink in its demand curve is because of how the other oligopolists react to changes in the firm's price. If the oligopoly decides to produce more and cut its price, the other members of the cartel will immediately match any price cuts—and therefore, a lower price brings very little increase in quantity sold.

If one firm cuts its price to \$300, it will be able to sell only 11,000 seats. However, if the airline seeks to raise prices, the other oligopolists will not raise their prices, and so the firm that raised prices will lose a considerable share of sales. For example, if the firm raises its price to \$550, its sales drop to 5,000 seats sold. Thus, if oligopolists always match price cuts by other firms in the cartel, but do not match price increases, then none of the

oligopolists will have a strong incentive to change prices, since the potential gains are minimal. This strategy can work like a silent form of cooperation, in which the cartel successfully manages to hold down output, increase price, and share a monopoly level of profits even without any legally enforceable agreement.

### A Kinked Demand Curve



Consider a member firm in an oligopoly cartel that is supposed to produce a quantity of 10,000 and sell at a price of \$500. The other members of the cartel can encourage this firm to honor its commitments by acting so that the firm faces a kinked demand curve. If the oligopolist attempts to expand output and reduce price slightly, other firms also cut prices immediately—so if the firm expands output to 11,000, the price per unit falls dramatically, to \$300. On the other side, if the oligopoly attempts to raise its price, other firms will not do so, so if the firm raises its price to \$550, its sales decline sharply to 5,000. Thus, the members of a cartel can discipline each other to stick to

the pre-agreed levels of quantity and price through a strategy of matching all price cuts but not matching any price increases.

Many real-world oligopolies, prodded by economic changes, legal and political pressures, and the egos of their top executives, go through episodes of cooperation and competition. If oligopolies could sustain cooperation with each other on output and pricing, they could earn profits as if they were a single monopoly. However, each firm in an oligopoly has an incentive to produce more and grab a bigger share of the overall market; when firms start behaving in this way, the market outcome in terms of prices and quantity can be similar to that of a highly competitive market.

## **Tradeoffs of Imperfect Competition**

Monopolistic competition is probably the single most common market structure in the U.S. economy. It provides powerful incentives for innovation, as firms seek to earn profits in the short run, while entry assures that firms do not earn economic profits in the long run. However, monopolistically competitive firms do not produce at the lowest point on their average cost curves. In addition, the endless search to impress consumers through product differentiation may lead to excessive social expenses on advertising and marketing.

Oligopoly is probably the second most common market structure. When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often buffeted by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior that only adds to the profits of a few large companies, with no corresponding benefit to consumers.

[Monopoly and Antitrust Policy](#) discusses the delicate judgments that go into this task.

**Note:**

**The Temptation to Defy the Law**

Oligopolistic firms have been called “cats in a bag,” as this chapter mentioned. The French detergent makers chose to “cozy up” with each other. The result? An uneasy and tenuous relationship. When the *Wall Street Journal* reported on the matter, it wrote: “According to a statement a Henkel manager made to the [French anti-trust] commission, the detergent makers wanted ‘to limit the intensity of the competition between them and clean up the market.’ Nevertheless, by the early 1990s, a price war had broken out among them.” During the soap executives’ meetings, which sometimes lasted more than four hours, complex pricing structures were established. “One [soap] executive recalled ‘chaotic’ meetings as each side tried to work out how the other had bent the rules.” Like many cartels, the soap cartel disintegrated due to the very strong temptation for each member to maximize its own individual profits.

How did this soap opera end? After an investigation, French antitrust authorities fined Colgate-Palmolive, Henkel, and Proctor & Gamble a total of €361 million (\$484 million). A similar fate befell the icemakers. Bagged ice is a commodity, a perfect substitute, generally sold in 7- or 22-pound bags. No one cares what label is on the bag. By agreeing to carve up the ice market, control broad geographic swaths of territory, and set prices, the icemakers moved from perfect competition to a monopoly model. After the agreements, each firm was the sole supplier of bagged ice to a region; there were profits in both the long run and the short run. According to the courts: “These companies illegally conspired to manipulate the marketplace.” Fines totaled about \$600,000—a steep fine considering a bag of ice sells for under \$3 in most parts of the United States.

Even though it is illegal in many parts of the world for firms to set prices and carve up a market, the temptation to earn higher profits makes it extremely tempting to defy the law.

## Key Concepts and Summary

An oligopoly is a situation where a few firms sell most or all of the goods in a market. Oligopolists earn their highest profits if they can band together as a cartel and act like a monopolist by reducing output and raising price. Since each member of the oligopoly can benefit individually from expanding output, such collusion often breaks down—especially since explicit collusion is illegal.

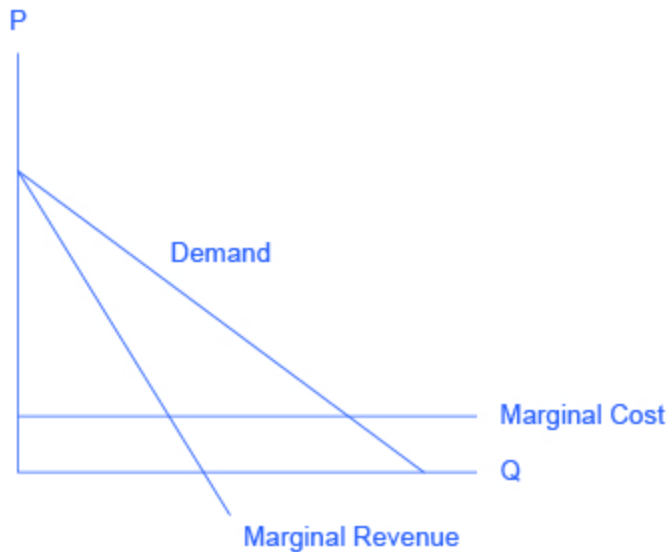
The prisoner's dilemma is an example of game theory. It shows how, in certain situations, all sides can benefit from cooperative behavior rather than self-interested behavior. However, the challenge for the parties is to find ways to encourage cooperative behavior.

## Self-Check Questions

### Exercise:

#### Problem:

Consider the curve shown in [\[link\]](#), which shows the market demand, marginal cost, and marginal revenue curve for firms in an oligopolistic industry. In this example, we assume firms have zero fixed costs.

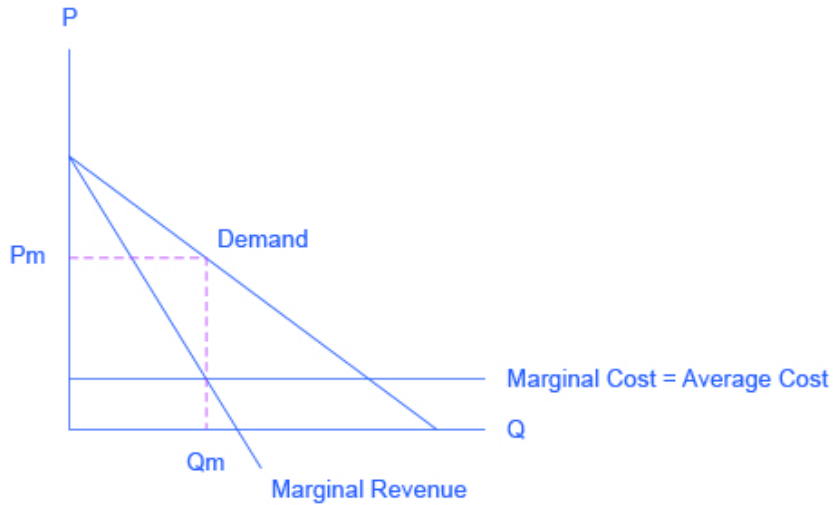


- a. Suppose the firms collude to form a cartel. What price will the cartel charge? What quantity will the cartel supply? How much profit will the cartel earn?
- b. Suppose now that the cartel breaks up and the oligopolistic firms compete as vigorously as possible by cutting the price and increasing sales. What will the industry quantity and price be? What will the collective profits be of all firms in the industry?
- c. Compare the equilibrium price, quantity, and profit for the cartel and cutthroat competition outcomes.

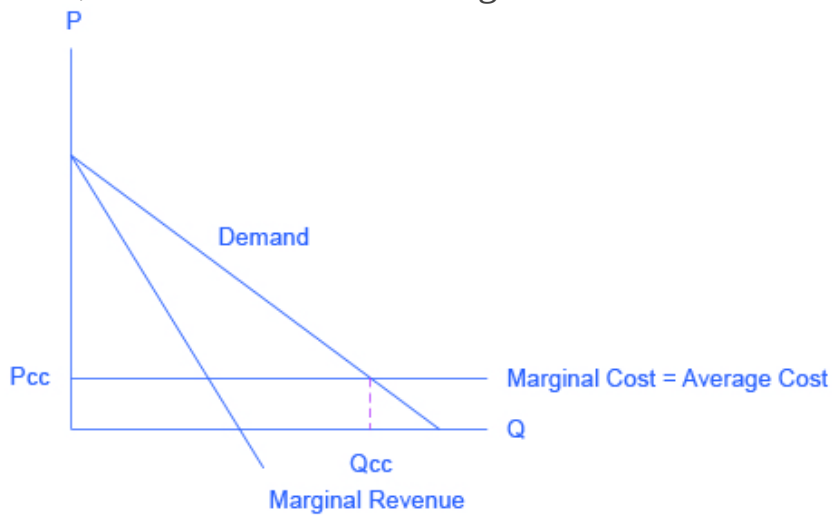
---

### Solution:

- a. If the firms form a cartel, they will act like a monopoly, choosing the quantity of output where  $MR = MC$ . Drawing a line from the monopoly quantity up to the demand curve shows the monopoly price. Assuming that fixed costs are zero, and with an understanding of cost and profit, we can infer that when the marginal cost curve is horizontal, average cost is the same as marginal cost. Thus, the cartel will earn positive economic profits equal to the area of the rectangle, with a base equal to the monopoly quantity and a height equal to the difference between price (on the demand above the monopoly quantity) and average cost, as shown in the following figure.



- b. The firms will expand output and cut price as long as there are profits remaining. The long-run equilibrium will occur at the point where average cost equals demand. As a result, the oligopoly will earn zero economic profits due to “cutthroat competition,” as shown in the next figure.



- c.  $P_c > P_{cc}$ .  $Q_c < Q_{cc}$ . Profit for the cartel is positive and large. Profit for cutthroat competition is zero.

**Exercise:**

**Problem:**

Sometimes oligopolies in the same industry are very different in size. Suppose we have a duopoly where one firm (Firm A) is large and the other firm (Firm B) is small, as shown in the prisoner's dilemma box in [\[link\]](#).

	Firm B colludes with Firm A	Firm B cheats by selling more output
Firm A colludes with Firm B	A gets \$1,000, B gets \$100	A gets \$800, B gets \$200
Firm A cheats by selling more output	A gets \$1,050, B gets \$50	A gets \$500, B gets \$20

Assuming that the payoffs are known to both firms, what is the likely outcome in this case?

---

**Solution:**

Firm B reasons that if it cheats and Firm A does not notice, it will double its money. Since Firm A's profits will decline substantially, however, it is likely that Firm A will notice and if so, Firm A will cheat also, with the result that Firm B will lose 90% of what it gained by cheating. Firm A will reason that Firm B is unlikely to risk cheating. If neither firm cheats, Firm A earns \$1000. If Firm A cheats, assuming Firm B does not cheat, A can boost its profits only a little, since Firm B is so small. If both firms cheat, then Firm A loses at least 50% of what it could have earned. The possibility of a small gain (\$50)

is probably not enough to induce Firm A to cheat, so in this case it is likely that both firms will collude.

## **Review Questions**

### **Exercise:**

#### **Problem:**

Will the firms in an oligopoly act more like a monopoly or more like competitors? Briefly explain.

### **Exercise:**

#### **Problem:**

Does each individual in a prisoner's dilemma benefit more from cooperation or from pursuing self-interest? Explain briefly.

### **Exercise:**

#### **Problem:**

What stops oligopolists from acting together as a monopolist and earning the highest possible level of profits?

## **Critical Thinking Questions**

### **Exercise:**

#### **Problem:**

Would you expect the kinked demand curve to be more extreme (like a right angle) or less extreme (like a normal demand curve) if each firm in the cartel produces a near-identical product like OPEC and petroleum? What if each firm produces a somewhat different product? Explain your reasoning.

### **Exercise:**

### Problem:

When OPEC raised the price of oil dramatically in the mid-1970s, experts said it was unlikely that the cartel could stay together over the long term—that the incentives for individual members to cheat would become too strong. More than forty years later, OPEC still exists. Why do you think OPEC has been able to beat the odds and continue to collude? *Hint:* You may wish to consider non-economic reasons.

## Problems

### Exercise:

#### Problem:

Mary and Raj are the only two growers who provide organically grown corn to a local grocery store. They know that if they cooperated and produced less corn, they could raise the price of the corn. If they work independently, they will each earn \$100. If they decide to work together and both lower their output, they can each earn \$150. If one person lowers output and the other does not, the person who lowers output will earn \$0 and the other person will capture the entire market and will earn \$200. [\[link\]](#) represents the choices available to Mary and Raj. What is the best choice for Raj if he is sure that Mary will cooperate? If Mary thinks Raj will cheat, what should Mary do and why? What is the prisoner's dilemma result? What is the preferred choice if they could ensure cooperation? A = Work independently; B = Cooperate and Lower Output. (Each results entry lists Raj's earnings first, and Mary's earnings second.)

	Mary	
	A	B
A		
B		

		A	B
Raj	A	(\$100, \$100)	(\$200, \$0)
	B	(\$0, \$200)	(\$150, \$150)

### Exercise:

#### Problem:

Jane and Bill are apprehended for a bank robbery. They are taken into separate rooms and questioned by the police about their involvement in the crime. The police tell them each that if they confess and turn the other person in, they will receive a lighter sentence. If they both confess, they will be each be sentenced to 30 years. If neither confesses, they will each receive a 20-year sentence. If only one confesses, the confessor will receive 15 years and the one who stayed silent will receive 35 years. [\[link\]](#) below represents the choices available to Jane and Bill. If Jane trusts Bill to stay silent, what should she do? If Jane thinks that Bill will confess, what should she do? Does Jane have a dominant strategy? Does Bill have a dominant strategy? A = Confess; B = Stay Silent. (Each results entry lists Jane's sentence first (in years), and Bill's sentence second.)

		Jane	
		A	B
Bill	A	(30, 30)	(15, 35)
	B	(35, 15)	(20, 20)

## References

The United States Department of Justice. "Antitrust Division." Accessed October 17, 2013. <http://www.justice.gov/atr/>.

eMarketer.com. 2014. "Total US Ad Spending to See Largest Increase Since 2004: Mobile advertising leads growth; will surpass radio, magazines and newspapers this year. Accessed March 12, 2015. <http://www.emarketer.com/Article/Total-US-Ad-Spending-See-Largest-Increase-Since-2004/1010982>.

Federal Trade Commission. "About the Federal Trade Commission." Accessed October 17, 2013. <http://www.ftc.gov/ftc/about.shtm>.

## Glossary

cartel

a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion

when firms act together to reduce output and keep prices high

duopoly

an oligopoly with only two firms

game theory

a branch of mathematics often used by economists that analyzes situations in which players must make decisions and then receive payoffs based on what decisions the other players make

kinked demand curve

a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

prisoner's dilemma

a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

## Introduction to Monopoly and Antitrust Policy

class="introduction"

### Oligopoly versus Competitors in the Marketplace

Large corporations, such as the natural gas producer Kinder Morgan, can bring economies of scale to the marketplace.

Will that benefit consumers?

Or is more competition better for consumers?

(Credit: modification of work by Derrick Coetzee/Flickr  
r Creative Commons)

**Note:****More than Cooking, Heating, and Cooling**

If you live in the United States, there is a slightly better than 50–50 chance your home is heated and cooled using natural gas. You may even use natural gas for cooking. However, those uses are not the primary uses of natural gas in the U.S. In 2012, according to the U.S. Energy Information Administration, home heating, cooling, and cooking accounted for just 18% of natural gas usage. What accounts for the rest? The greatest uses for natural gas are the generation of electric power (39%) and in industry (30%). Together these three uses for natural gas touch many areas of our lives, so why would there be any opposition to a merger of two natural gas firms? After all, a merger could mean increased efficiencies and reduced costs to people like you and me.

In October 2011, Kinder Morgan and El Paso Corporation, two natural gas firms, announced they were merging. The announcement stated the combined firm would link “nearly every major production region with markets,” cut costs by “eliminating duplication in pipelines and other assets,” and that “the savings could be passed on to consumers.”

The objection? The \$21.1 billion deal would give Kinder Morgan control of more than 80,000 miles of pipeline, making the new firm the third largest energy producer in North America. As the third largest energy producer, policymakers and the public wondered whether the cost savings

really would be passed on to consumers, or would the merger give Kinder Morgan a strong oligopoly position in the natural gas marketplace?

That brings us to the central question this chapter poses: What should the balance be between corporate size and a larger number of competitors in a marketplace? We will also consider what role the government should play in this balancing act.

**Note:**

**Introduction to Monopoly and Antitrust Policy**

In this chapter, you will learn about:

- Corporate Mergers
- Regulating Anticompetitive Behavior
- Regulating Natural Monopolies
- The Great Deregulation Experiment

The previous chapters on the theory of the firm identified three important lessons: First, that competition, by providing consumers with lower prices and a variety of innovative products, is a good thing; second, that large-scale production can dramatically lower average costs; and third, that markets in the real world are rarely perfectly competitive. As a consequence, government policymakers must determine how much to intervene to balance the potential benefits of large-scale production against the potential loss of competition that can occur when businesses grow in size, especially through mergers.

For example, in 2011, AT&T and T-Mobile proposed a merger. At the time, there were only four major mobile phone service providers. The proposal was blocked by both the Justice Department and the FCC.

The two companies argued that the merger would benefit consumers, who would be able to purchase better telecommunications services at a cheaper price because the newly created firm would be able to produce more

efficiently by taking advantage of economies of scale and eliminating duplicate investments. However, a number of activist groups like the Consumer Federation of America and Public Knowledge expressed fears that the merger would reduce competition and lead to higher prices for consumers for decades to come. In December 2006, the federal government allowed the merger to proceed. By 2009, the new post-merger AT&T was the eighth largest company by revenues in the United States, and by that measure the largest telecommunications company in the world. Economists have spent – and will still spend – years trying to determine whether the merger of AT&T and BellSouth, as well as other smaller mergers of telecommunications companies at about this same time, helped consumers, hurt them, or did not make much difference.

This chapter discusses public policy issues about competition. How can economists and governments determine when mergers of large companies like AT&T and BellSouth should be allowed and when they should be blocked? The government also plays a role in policing anticompetitive behavior other than mergers, like prohibiting certain kinds of contracts that might restrict competition. In the case of natural monopoly, however, trying to preserve competition probably will not work very well, and so government will often resort to regulation of price and/or quantity of output. In recent decades, there has been a global trend toward less government intervention in the price and output decisions of businesses.

## Corporate Mergers

By the end of this section, you will be able to:

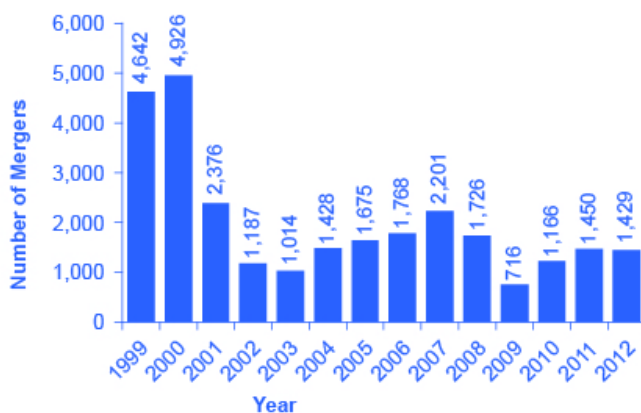
- Explain antitrust law and its significance
- Calculate concentration ratios
- Calculate the Herfindahl-Herschman Index (HHI)
- Evaluate methods of antitrust regulation

A corporate **merger** occurs when two formerly separate firms combine to become a single firm. When one firm purchases another, it is called an **acquisition**. An acquisition may not look just like a merger, since the newly purchased firm may continue to be operated under its former company name. Mergers can also be lateral, where two firms of similar sizes combine to become one. However, both mergers and acquisitions lead to two formerly separate firms being under common ownership, and so they are commonly grouped together.

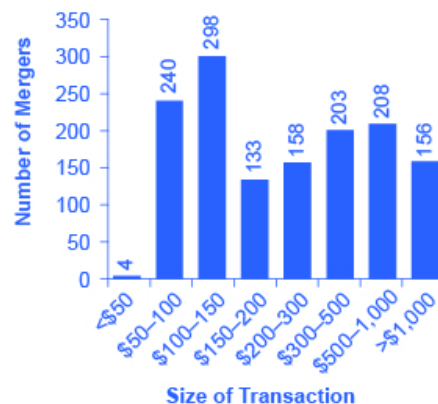
## Regulations for Approving Mergers

Since a merger combines two firms into one, it can reduce the extent of competition between firms. Therefore, when two U.S. firms announce a merger or acquisition where at least one of the firms is above a minimum size of sales (a threshold that moves up gradually over time, and was at \$70.9 million in 2013), or certain other conditions are met, they are required under law to notify the U.S. Federal Trade Commission (FTC). The left-hand panel of [\[link\]](#) (a) shows the number of mergers submitted for review to the FTC each year from 1999 to 2012. Mergers were very high in the late 1990s, diminished in the early 2000s, and then rebounded somewhat in a cyclical fashion. The right-hand panel of [\[link\]](#) (b) shows the distribution of those mergers submitted for review in 2012 as measured by the size of the transaction. It is important to remember that this total leaves out many small mergers under \$50 million, which only need to be reported in certain limited circumstances. About a quarter of all reported merger and acquisition transactions in 2012 exceeded \$500 million, while about 11 percent exceeded \$1 billion. In 2014, the FTC took action against mergers likely to stifle competition in markets worth 18.6 billion in sales.

## Number and Size of Mergers



(a) Number of mergers submitted for review by the Federal Trade Commission, 1999-2012



(b) Size of transaction for mergers submitted for review in 2012 (in millions of dollars)

(a) The number of mergers in 1999 and 2000 were relatively high compared to the annual numbers seen from 2001–2012. While 2001 and 2007 saw a high number of mergers, these were still only about half the number of mergers in 1999 and 2000. (b) In 2012, the greatest number of mergers submitted for review was for transactions between \$100 and \$150 million.

The laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones, are called **antitrust laws**. Before a large merger happens, the antitrust regulators at the FTC and the U.S. Department of Justice can allow the merger, prohibit it, or allow it if certain conditions are met. One common condition is that the merger will be allowed if the firm agrees to sell off certain parts. For example, in 2006, Johnson & Johnson bought the Pfizer’s “consumer health” division, which included well-known brands like Listerine mouthwash and Sudafed cold medicine. As a condition of allowing the merger, Johnson & Johnson was required to sell off six brands to other firms, including Zantac® heartburn relief medication, Cortizone anti-itch cream, and Balmex diaper rash medication, to preserve a greater degree of competition in these markets.

The U.S. government approves most proposed mergers. In a market-oriented economy, firms have the freedom to make their own choices. Private firms generally have the freedom to:

- expand or reduce production
- set the price they choose
- open new factories or sales facilities or close them
- hire workers or to lay them off
- start selling new products or stop selling existing ones

If the owners want to acquire a firm or be acquired, or to merge with another firm, this decision is just one of many that firms are free to make. In these conditions, the managers of private firms will sometimes make mistakes. They may close down a factory which, it later turns out, would have been profitable. They may start selling a product that ends up losing money. A merger between two companies can sometimes lead to a clash of corporate personalities that makes both firms worse off. But the fundamental belief behind a market-oriented economy is that firms, not governments, are in the best position to know if their actions will lead to attracting more customers or producing more efficiently.

Indeed, government regulators agree that most mergers are beneficial to consumers. As the Federal Trade Commission has noted on its website (as of November, 2013): “Most mergers actually benefit competition and consumers by allowing firms to operate more efficiently.” At the same time, the FTC recognizes, “Some [mergers] are likely to lessen competition. That, in turn, can lead to higher prices, reduced availability of goods or services, lower quality of products, and less innovation. Indeed, some mergers create a concentrated market, while others enable a single firm to raise prices.” The challenge for the antitrust regulators at the FTC and the U.S. Department of Justice is to figure out when a merger may hinder competition. This decision involves both numerical tools and some judgments that are difficult to quantify. The following Clear it Up helps explain how antitrust laws came about.

**Note:**

### What is U.S. antitrust law?

In the closing decades of the 1800s, many industries in the U.S. economy were dominated by a single firm that had most of the sales for the entire country. Supporters of these large firms argued that they could take advantage of economies of scale and careful planning to provide consumers with products at low prices. However, critics pointed out that when competition was reduced, these firms were free to charge more and make permanently higher profits, and that without the goading of competition, it was not clear that they were as efficient or innovative as they could be.

In many cases, these large firms were organized in the legal form of a “trust,” in which a group of formerly independent firms were consolidated together by mergers and purchases, and a group of “trustees” then ran the companies as if they were a single firm. Thus, when the U.S. government passed the **Sherman Antitrust Act** in 1890 to limit the power of these trusts, it was called an antitrust law. In an early demonstration of the law’s power, the U.S. Supreme Court in 1911 upheld the government’s right to break up Standard Oil, which had controlled about 90% of the country’s oil refining, into 34 independent firms, including Exxon, Mobil, Amoco, and Chevron. In 1914, the **Clayton Antitrust Act** outlawed mergers and acquisitions (where the outcome would be to “substantially lessen competition” in an industry), price discrimination (where different customers are charged different prices for the same product), and tied sales (where purchase of one product commits the buyer to purchase some other product). Also in 1914, the Federal Trade Commission (FTC) was created to define more specifically what competition was unfair. In 1950, the **Celler-Kefauver Act** extended the Clayton Act by restricting vertical and conglomerate mergers. In the twenty-first century, the FTC and the U.S. Department of Justice continue to enforce antitrust laws.

### The Four-Firm Concentration Ratio

Regulators have struggled for decades to measure the degree of monopoly power in an industry. An early tool was the **concentration ratio**, which measures what share of the total sales in the industry are accounted for by

the largest firms, typically the top four to eight firms. For an explanation of how high market concentrations can create inefficiencies in an economy, refer to [Monopoly](#).

Say that the market for replacing broken automobile windshields in a certain city has 18 firms with the market shares shown in [\[link\]](#), where the **market share** is each firm’s proportion of total sales in that market. The four-firm concentration ratio is calculated by adding the market shares of the four largest firms: in this case,  $16 + 10 + 8 + 6 = 40$ . This concentration ratio would not be considered especially high, because the largest four firms have less than half the market.

If the market shares in the market for replacing automobile windshields are:	
Smooth as Glass Repair Company	16% of the market
The Auto Glass Doctor Company	10% of the market
Your Car Shield Company	8% of the market
Seven firms that each have 6% of the market	42% of the market, combined
Eight firms that each have 3% of the market	24% of the market, combined
Then the four-firm concentration ratio is $16 + 10 + 8 + 6 = 40$ .	

Calculating Concentration Ratios from Market Shares

The concentration ratio approach can help to clarify some of the fuzziness over deciding when a merger might affect competition. For instance, if two

of the smallest firms in the hypothetical market for repairing automobile windshields merged, the four-firm concentration ratio would not change—which implies that there is not much worry that the degree of competition in the market has notably diminished. However, if the top two firms merged, then the four-firm concentration ratio would become 46 (that is,  $26 + 8 + 6$ ). While this concentration ratio is modestly higher, the four-firm concentration ratio would still be less than half, so such a proposed merger might barely raise an eyebrow among antitrust regulators.

**Note:**

Visit this [website](#) to read an article about Google’s run-in with the FTC.



## The Herfindahl-Hirschman Index

A four-firm concentration ratio is a simple tool, which may reveal only part of the story. For example, consider two industries that both have a four-firm concentration ratio of 80. However, in one industry five firms each control 20% of the market, while in the other industry, the top firm holds 77% of the market and all the other firms have 1% each. Although the four-firm concentration ratios are identical, it would be reasonable to worry more about the extent of competition in the second case—where the largest firm is nearly a monopoly—than in the first.

Another approach to measuring industry concentration that can distinguish between these two cases is called the **Herfindahl-Hirschman Index (HHI)**. The HHI, as it is often called, is calculated by summing the squares

of the market share of each firm in the industry, as the following Work it Out shows.

**Note:**

**Calculating HHI**

Step 1. Calculate the HHI for a monopoly with a market share of 100%. Because there is only one firm, it has 100% market share. The HHI is  $100^2 = 10,000$ .

Step 2. For an extremely competitive industry, with dozens or hundreds of extremely small competitors, the value of the HHI might drop as low as 100 or even less. Calculate the HHI for an industry with 100 firms that each have 1% of the market. In this case, the HHI is  $100(1^2) = 100$ .

Step 3. Calculate the HHI for the industry shown in [\[link\]](#). In this case, the HHI is  $16^2 + 10^2 + 8^2 + 7(6^2) + 8(3^2) = 744$ .

Step 4. Note that the HHI gives greater weight to large firms.

Step 5. Consider the example given earlier, comparing one industry where five firms each have 20% of the market with an industry where one firm has 77% and the other 23 firms have 1% each. The two industries have the same four-firm concentration ratio of 80. But the HHI for the first industry is  $5(20^2) = 2,000$ , while the HHI for the second industry is much higher at  $77^2 + 23(1^2) = 5,952$ .

Step 6. Note that the near-monopolist in the second industry drives up the HHI measure of industrial concentration.

Step 7. Review [\[link\]](#) which gives some examples of the four-firm concentration ratio and the HHI in various U.S. industries in 2009. (You can find market share data from multiple industry sources. Data in the table are from: Verizon (for wireless), *The Wall Street Journal* (for automobiles), IDC Worldwide (for computers) and the U.S. Bureau of Transportation Statistics (for airlines).)

<b>U.S. Industry</b>	<b>Four-Firm Ratio</b>	<b>HHI</b>
<i>Wireless</i>	91	2,311
Largest five: Verizon, AT&T, Sprint, T-Mobile, MetroPCS		
<i>Automobiles</i>	63	1,121
Largest five: GM, Toyota, Ford, Honda, Chrysler		
<i>Computers</i>	74	1,737
Largest five: HP, Dell, Acer, Apple, Toshiba		
<i>Airlines</i>	44	536
Largest five: Southwest, American, Delta, United, U.S. Airways		
Examples of Concentration Ratios and HHIs in the U.S. Economy, 2009		

In the 1980s, the FTC followed these guidelines: If a merger would result in an HHI of less than 1,000, the FTC would probably approve it. If a merger would result in an HHI of more than 1,800, the FTC would probably challenge it. If a merger would result in an HHI between 1,000 and 1,800, then the FTC would scrutinize the plan and make a case-by-case decision. However, in the last several decades, the antitrust enforcement authorities have moved away from relying as heavily on measures of concentration ratios and HHIs to determine whether a merger will be allowed, and instead

carried out more case-by-case analysis on the extent of competition in different industries.

## **New Directions for Antitrust**

Both the four-firm concentration ratio and the Herfindahl-Hirschman index share some weaknesses. First, they begin from the assumption that the “market” under discussion is well-defined, and the only question is measuring how sales are divided in that market. Second, they are based on an implicit assumption that competitive conditions across industries are similar enough that a broad measure of concentration in the market is enough to make a decision about the effects of a merger. These assumptions, however, are not always correct. In response to these two problems, the antitrust regulators have been changing their approach in the last decade or two.

Defining a **market** is often controversial. For example, Microsoft in the early 2000s had a dominant share of the software for computer operating systems. However, in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. A narrowly defined market will tend to make concentration appear higher, while a broadly defined market will tend to make it appear smaller.

There are two especially important shifts affecting how markets are defined in recent decades: one centers on technology and the other centers on globalization. In addition, these two shifts are interconnected. With the vast improvement in communications technologies, including the development of the Internet, a consumer can order books or pet supplies from all over the country or the world. As a result, the degree of competition many local retail businesses face has increased. The same effect may operate even more strongly in markets for business supplies, where so-called “business-to-business” websites can allow buyers and suppliers from anywhere in the world to find each other.

Globalization has changed the boundaries of markets. As recently as the 1970s, it was common for measurements of concentration ratios and HHIs

to stop at national borders. Now, many industries find that their competition comes from the global market. A few decades ago, three companies, General Motors, Ford, and Chrysler, dominated the U.S. auto market. By 2014, however, these three firms were making less than half of U.S. auto sales, and facing competition from well-known car manufacturers such as Toyota, Honda, Nissan, Volkswagen, Mitsubishi, and Mazda. When HHIs are calculated with a global perspective, concentration in most major industries—including cars—is lower than in a purely domestic context.

Because attempting to define a particular market can be difficult and controversial, the Federal Trade Commission has begun to look less at market share and more at the data on actual competition between businesses. For example, in February 2007, Whole Foods Market and Wild Oats Market announced that they wished to merge. These were the two largest companies in the market that the government defined as “premium natural and organic supermarket chains.” However, one could also argue that they were two relatively small companies in the broader market for all stores that sell groceries or specialty food products.

Rather than relying on a market definition, the government antitrust regulators looked at detailed evidence on profits and prices for specific stores in different cities, both before and after other competitive stores entered or exited. Based on that evidence, the Federal Trade Commission decided to block the merger. After two years of legal battles, the merger was eventually allowed in 2009 under the conditions that Whole Foods sell off the Wild Oats brand name and a number of individual stores, to preserve competition in certain local markets. For more on the difficulties of defining markets, refer to [Monopoly](#).

This new approach to antitrust regulation involves detailed analysis of specific markets and companies, instead of defining a market and counting up total sales. A common starting point is for antitrust regulators to use statistical tools and real-world evidence to estimate the **demand curves** and **supply curves** faced by the firms that are proposing the merger. A second step is to specify how competition occurs in this specific industry. Some possibilities include competing to cut prices, to raise output, to build a brand name through advertising, and to build a reputation for good service

or high quality. With these pieces of the puzzle in place, it is then possible to build a statistical model that estimates the likely outcome for consumers if the two firms are allowed to merge. Of course, these models do require some degree of subjective judgment, and so they can become the subject of legal disputes between the antitrust authorities and the companies that wish to merge.

## **Key Concepts and Summary**

A corporate merger involves two private firms joining together. An acquisition refers to one firm buying another firm. In either case, two formerly independent firms become one firm. Antitrust laws seek to ensure active competition in markets, sometimes by preventing large firms from forming through mergers and acquisitions, sometimes by regulating business practices that might restrict competition, and sometimes by breaking up large firms into smaller competitors.

A four-firm concentration ratio is one way of measuring the extent of competition in a market. It is calculated by adding the market shares—that is, the percentage of total sales—of the four largest firms in the market. A Herfindahl-Hirschman Index (HHI) is another way of measuring the extent of competition in a market. It is calculated by taking the market shares of all firms in the market, squaring them, and then summing the total.

The forces of globalization and new communications and information technology have increased the level of competition faced by many firms by increasing the amount of competition from other regions and countries.

## **Self-Check Questions**

### **Exercise:**

#### **Problem:**

Is it true that both the four-firm concentration ratio and the Herfindahl-Hirschman Index can be affected by a merger between two firms that are not already in the top four by size? Explain briefly.

---

**Solution:**

Yes, it is true. The HHI example is easy enough: since the market shares of all firms are included in the HHI calculation, a merger between two of the firms will change the HHI. For the four-firm concentration ratio, it is quite possible that a merger between, say, the fifth and sixth largest firms in the market could create a new firm that is then ranked in the top four in the market. In this case, a merger of two firms, neither in the top four, would still change the four-firm concentration ratio.

**Exercise:****Problem:**

Is it true that the four-firm concentration ratio puts more emphasis on one or two very large firms, while the Herfindahl-Hirshman Index puts more emphasis on all the firms in the entire market? Explain briefly.

---

**Solution:**

No, it is not true. The HHI includes the market shares of all firms in its calculation, but the squaring of the market shares has the effect of making the impact of the largest firms relatively bigger than in the 4-firm or 8-firm ratio.

**Exercise:****Problem:**

Some years ago, two intercity bus companies, Greyhound Lines, Inc. and Trailways Transportation System, wanted to merge. One possible definition of the market in this case was “the market for intercity bus service.” Another possible definition was “the market for intercity transportation, including personal cars, car rentals, passenger trains, and commuter air flights.” Which definition do you think the bus companies preferred, and why?

---

**Solution:**

The bus companies wanted the broader market definition (i.e., the second definition). If the narrow definition had been used, the combined bus companies would have had a near-monopoly on the market for intercity bus service. But they had only a sliver of the market for intercity transportation when everything else was included. The merger was allowed.

**Exercise:**

**Problem:**

As a result of globalization and new information and communications technology, would you expect that the definitions of markets used by antitrust authorities will become broader or narrower?

---

**Solution:**

The common expectation is that the definition of markets will become broader because of greater competition from faraway places. However, this broadening doesn't necessarily mean that antitrust authorities can relax. There is also a fear that companies with a local or national monopoly may use the new opportunities to extend their reach across national borders, and that it will be difficult for national authorities to respond.

## **Review Questions**

**Exercise:**

**Problem:** What is a corporate merger? What is an acquisition?

**Exercise:**

**Problem:** What is the goal of antitrust policies?

**Exercise:**

**Problem:**

How is a four-firm concentration ratio measured? What does a high measure mean about the extent of competition?

**Exercise:****Problem:**

How is a Herfindahl-Hirshman Index measured? What does a low measure mean about the extent of competition?

**Exercise:****Problem:**

Why can it be difficult to decide what a “market” is for purposes of measuring competition?

**Critical Thinking Questions****Exercise:****Problem:**

Does either the four-firm concentration ratio or the HHI directly measure the amount of competition in an industry? Why or why not?

**Exercise:****Problem:**

What would be evidence of serious competition between firms in an industry? Can you identify two highly competitive industries?

**Problems****Exercise:**

**Problem:**

Use [\[link\]](#) to calculate the four-firm concentration ratio for the U.S. auto market. Does this indicate a concentrated market or not?

GM	19%
Ford	17%
Toyota	14%
Chrysler	11%

Global Auto Manufacturers with Top Four U.S. Market Share, June 2013(Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

**Exercise:****Problem:**

Use [\[link\]](#) and [\[link\]](#) to calculate the Herfindal-Hirschman Index for the U.S. auto market. Would the FTC approve a merger between GM and Ford?

Honda	10%
Nissan	7%

Hyundai	5%
Kia	4%
Subaru	3%
Volkswagen	3%

Global Auto Manufacturers with additional U.S. Market Share, June 2013(Source: <http://www.zacks.com/commentary/27690/auto-industry-stock-outlook-june-2013>)

## Glossary

acquisition

when one firm purchases another

antitrust laws

laws that give government the power to block certain mergers, and even in some cases to break up large firms into smaller ones

concentration ratio

an early tool to measure the degree of monopoly power in an industry; measures what share of the total sales in the industry are accounted for by the largest firms, typically the top four to eight firms

four-firm concentration ratio

the percentage of the total sales in the industry that are accounted for by the largest four firms

Herfindahl-Hirschman Index (HHI)

approach to measuring market concentration by adding the square of the market share of each firm in the industry

market share

the percentage of total sales in the market

merger

when two formerly separate firms combine to become a single firm

## Regulating Anticompetitive Behavior

By the end of this section, you will be able to:

- Analyze restrictive practices
- Explain tying sales, bundling, and predatory pricing
- Evaluate a real-world situation of possible anticompetitive and restrictive practices

The U.S. antitrust laws reach beyond blocking mergers that would reduce competition to include a wide array of anticompetitive practices. For example, it is illegal for competitors to form a cartel to collude to make pricing and output decisions, as if they were a monopoly firm. The Federal Trade Commission and the U.S. Department of Justice prohibit firms from agreeing to fix prices or output, rigging bids, or sharing or dividing markets by allocating customers, suppliers, territories, or lines of commerce.

In the late 1990s, for example, the antitrust regulators prosecuted an international cartel of vitamin manufacturers, including the Swiss firm Hoffman-La Roche, the German firm BASF, and the French firm Rhone-Poulenc. These firms reached agreements on how much to produce, how much to charge, and which firm would sell to which customers. The high-priced vitamins were then bought by firms like General Mills, Kellogg, Purina-Mills, and Proctor and Gamble, which pushed up the prices more. Hoffman-La Roche pleaded guilty in May 1999 and agreed both to pay a fine of \$500 million and to have at least one top executive serve four months of jail time.

Under U.S. antitrust laws, monopoly itself is not illegal. If a firm has a monopoly because of a newly patented invention, for example, the law explicitly allows a firm to earn higher-than-normal profits for a time as a reward for innovation. If a firm achieves a large share of the market by producing a better product at a lower price, such behavior is not prohibited by antitrust law.

## Restrictive Practices

Antitrust law includes rules against **restrictive practices**—practices that do not involve outright agreements to raise price or to reduce the quantity produced, but that might have the effect of reducing competition. Antitrust cases involving restrictive practices are often controversial, because they delve into specific contracts or agreements between firms that are allowed in some cases but not in others.

For example, if a product manufacturer is selling to a group of dealers who then sell to the general public it is illegal for the manufacturer to demand a **minimum resale price maintenance agreement**, which would require the dealers to sell for at least a certain minimum price. A minimum price contract is illegal because it would restrict competition among dealers. However, the manufacturer is legally allowed to “suggest” minimum prices and to stop selling to dealers who regularly undercut the suggested price. If you think this rule sounds like a fairly subtle distinction, you are right.

An **exclusive dealing** agreement between a manufacturer and a dealer can be legal or illegal. It is legal if the purpose of the contract is to encourage competition between dealers. For example, it is legal for the Ford Motor Company to sell its cars to only Ford dealers, for General Motors to sell to only GM dealers, and so on. However, exclusive deals may also limit competition. If one large retailer obtained the exclusive rights to be the sole distributor of televisions, computers, and audio equipment made by a number of companies, then this exclusive contract would have an anticompetitive effect on other retailers.

**Tying sales** happen when a customer is required to buy one product only if the customer also buys a second product. Tying sales are controversial because they force consumers to purchase a product that they may not actually want or need. Further, the additional, required products are not necessarily advantageous to the customer. Suppose that to purchase a popular DVD, the store required that you also purchase a portable TV of a certain model. These products are only loosely related, thus there is no reason to make the purchase of one contingent on the other. Even if a customer was interested in a portable TV, the tying to a particular model prevents the customer from having the option of selecting one from the numerous types available in the market. A related, but not identical, concept

is called **bundling**, where two or more products are sold as one. Bundling typically offers an advantage for the consumer by allowing them to acquire multiple products or services for a better price. For example, several cable companies allow customers to buy products like cable, internet, and a phone line through a special price available through bundling. Customers are also welcome to purchase these products separately, but the price of bundling is usually more appealing.

In some cases, tying sales and bundling can be viewed as anticompetitive. However, in other cases they may be legal and even common. It is common for people to purchase season tickets to a sports team or a set of concerts so that they can be guaranteed tickets to the few contests or shows that are most popular and likely to sell out. Computer software manufacturers may often bundle together a number of different programs, even when the buyer wants only a few of the programs. Think about the software that is included in a new computer purchase, for example.

Recall from the chapter on [Monopoly](#) that predatory pricing occurs when the existing firm (or firms) reacts to a new firm by dropping prices very low, until the new firm is driven out of the market, at which point the existing firm raises prices again. This pattern of pricing is aimed at deterring the entry of new firms into the market. But in practice, it can be hard to figure out when pricing should be considered predatory. Say that American Airlines is flying between two cities, and a new airline starts flying between the same two cities, at a lower price. If American Airlines cuts its price to match the new entrant, is this predatory pricing? Or is it just market competition at work? A commonly proposed rule is that if a firm is selling for less than its average variable cost—that is, at a price where it should be shutting down—then there is evidence for predatory pricing. But calculating in the real world what costs are variable and what costs are fixed is often not obvious, either.

The Microsoft antitrust case embodies many of these gray areas in restrictive practices, as the next Clear it Up shows.

**Note:**

Did Microsoft® engage in anticompetitive and restrictive practices?

The most famous restrictive practices case of recent years was a series of lawsuits by the U.S. government against Microsoft—lawsuits that were encouraged by some of Microsoft’s competitors. All sides admitted that Microsoft’s Windows program had a near-monopoly position in the market for the software used in general computer operating systems. All sides agreed that the software had many satisfied customers. All sides agreed that the capabilities of computer software that was compatible with Windows—both software produced by Microsoft and that produced by other companies—had expanded dramatically in the 1990s. Having a **monopoly** or a near-monopoly is not necessarily illegal in and of itself, but in cases where one company controls a great deal of the market, antitrust regulators look at any allegations of restrictive practices with special care. The antitrust regulators argued that Microsoft had gone beyond profiting from its software innovations and its dominant position in the software market for operating systems, and had tried to use its market power in operating systems software to take over other parts of the software industry. For example, the government argued that Microsoft had engaged in an anticompetitive form of exclusive dealing by threatening computer makers that, if they did not leave another firm’s software off their machines (specifically, Netscape’s Internet browser), then Microsoft would not sell them its operating system software. Microsoft was accused by the government antitrust regulators of tying together its Windows operating system software, where it had a monopoly, with its Internet Explorer browser software, where it did not have a monopoly, and thus using this bundling as an anticompetitive tool. Microsoft was also accused of a form of predatory pricing; namely, giving away certain additional software products for free as part of Windows, as a way of driving out the competition from other makers of software.

In April 2000, a federal court held that Microsoft’s behavior had crossed the line into unfair competition, and recommended that the company be broken into two competing firms. However, that penalty was overturned on appeal, and in November 2002 Microsoft reached a settlement with the government that it would end its restrictive practices.

The concept of restrictive practices is continually evolving, as firms seek new ways to earn profits and government regulators define what is permissible and what is not. A situation where the law is evolving and changing is always somewhat troublesome, since laws are most useful and fair when firms know what they are in advance. In addition, since the law is open to interpretation, competitors who are losing out in the market can accuse successful firms of anticompetitive restrictive practices, and try to win through government regulation what they have failed to accomplish in the market. Officials at the Federal Trade Commission and the Department of Justice are, of course, aware of these issues, but there is no easy way to resolve them.

## **Key Concepts and Summary**

Firms are blocked by antitrust authorities from openly colluding to form a cartel that will reduce output and raise prices. Companies sometimes attempt to find other ways around these restrictions and, consequently, many antitrust cases involve restrictive practices that can reduce competition in certain circumstances, like tie-in sales, bundling, and predatory pricing.

## **Self-Check Question**

### **Exercise:**

#### **Problem:**

Why would a firm choose to use one or more of the anticompetitive practices described in [Regulating Anticompetitive Behavior](#)?

---

#### **Solution:**

Because outright collusion to raise profits is illegal and because existing regulations include gray areas which firms may be able to exploit.

## **Review Questions**

### **Exercise:**

#### **Problem:**

What is a minimum resale price maintenance agreement? How might it reduce competition and when might it be acceptable?

### **Exercise:**

#### **Problem:**

What is exclusive dealing? How might it reduce competition and when might it be acceptable?

### **Exercise:**

#### **Problem:**

What is a tie-in sale? How might it reduce competition and when might it be acceptable?

### **Exercise:**

#### **Problem:**

What is predatory pricing? How might it reduce competition, and why might it be difficult to tell when it should be illegal?

## **Critical Thinking Questions**

### **Exercise:**

#### **Problem:**

Can you think of any examples of successful predatory pricing in the real world?

### **Exercise:**

**Problem:**

If you were developing a product (like a web browser) for a market with significant barriers to entry, how would you try to get your product into the market successfully?

**Glossary****bundling**

a situation in which multiple products are sold as one

**exclusive dealing**

an agreement that a dealer will sell only products from one manufacturer

**minimum resale price maintenance agreement**

an agreement that requires a dealer who buys from a manufacturer to sell for at least a certain minimum price

**restrictive practices**

practices that reduce competition but that do not involve outright agreements between firms to raise prices or to reduce the quantity produced

**tying sales**

a situation where a customer is allowed to buy one product only if the customer also buys another product

## Regulating Natural Monopolies

By the end of this section, you will be able to:

- Evaluate the appropriate competition policy for a natural monopoly
- Interpret a graph of regulatory choices
- Contrast cost-plus and price cap regulation

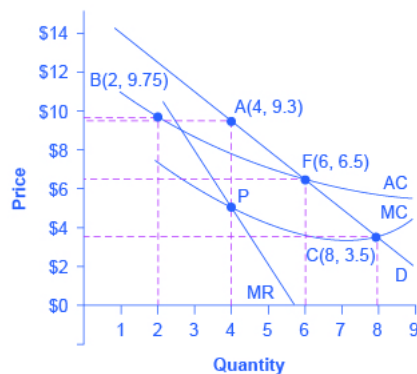
Most true monopolies today in the U.S. are regulated, natural monopolies. A natural monopoly poses a difficult challenge for competition policy, because the structure of costs and demand seems to make competition unlikely or costly. A **natural monopoly** arises when average costs are declining over the range of production that satisfies market demand. This typically happens when fixed costs are large relative to variable costs. As a result, one firm is able to supply the total quantity demanded in the market at lower cost than two or more firms—so splitting up the natural monopoly would raise the average cost of production and force customers to pay more.

Public utilities, the companies that have traditionally provided water and electrical service across much of the United States, are leading examples of natural monopoly. It would make little sense to argue that a local water company should be broken up into several competing companies, each with its own separate set of pipes and water supplies. Installing four or five identical sets of pipes under a city, one for each water company, so that each household could choose its own water provider, would be terribly costly. The same argument applies to the idea of having many competing companies for delivering electricity to homes, each with its own set of wires. Before the advent of wireless phones, the argument also applied to the idea of many different phone companies, each with its own set of phone wires running through the neighborhood.

### The Choices in Regulating a Natural Monopoly

So what then is the appropriate competition policy for a natural monopoly? [\[link\]](#) illustrates the case of natural monopoly, with a market demand curve that cuts through the downward-sloping portion of the **average cost curve**. Points A, B, C, and F illustrate four of the main choices for regulation. [\[link\]](#) outlines the regulatory choices for dealing with a natural monopoly.

Regulatory Choices in Dealing with Natural Monopoly



A natural monopoly will maximize profits by producing at the quantity where marginal revenue (MR) equals marginal costs (MC) and by then looking to the market demand curve to see what price to charge for this quantity. This monopoly will produce at point A, with a quantity of 4 and a price of 9.3. If antitrust regulators split this company exactly in half, then each half would produce at point B, with average costs of 9.75 and output of 2. The

regulators might require the firm to produce where marginal cost crosses the market demand curve at point C. However, if the firm is required to produce at a quantity of 8 and sell at a price of 3.5, the firm will suffer from losses. The most likely choice is point F, where the firm is required to produce a quantity of 6 and charge a price of 6.5.

Quantity	Price	Total Revenue*	Marginal Revenue	Total Cost	Marginal Cost	Average Cost
1	14.7	14.7	-	11.0	-	11.00
2	12.4	24.7	10.0	19.5	8.5	9.75
3	10.6	31.7	7.0	25.5	6.0	8.50
4	9.3	37.2	5.5	31.0	5.5	7.75
5	8.0	40.0	2.8	35.0	4.0	7.00
6	6.5	39.0	-1.0	39.0	4.0	6.50
7	5.0	35.0	-4.0	42.0	3.0	6.00
8	3.5	28.0	-7.0	45.5	3.5	5.70
9	2.0	18.0	-10.0	49.5	4.0	5.5

Regulatory Choices in Dealing with Natural Monopoly(\*Total Revenue is given by multiplying price and quantity. However, some of the price values in this table have been rounded for ease of presentation.)

The first possibility is to leave the natural monopoly alone. In this case, the monopoly will follow its normal approach to maximizing profits. It determines the quantity where  $MR = MC$ , which happens at point P at a quantity of 4. The firm then looks to point A on the demand curve to find that it can charge a price of 9.3 for that profit-maximizing quantity. Since the price is above the average cost curve, the natural monopoly would earn economic profits.

A second outcome arises if antitrust authorities decide to divide the company, so that the new firms can compete. As a simple example, imagine that the company is cut in half. Thus, instead of one large firm producing a quantity of 4, two half-size firms each produce a quantity of 2. Because of the declining average cost curve (AC), the average cost of production for each of the half-size companies each producing 2, as shown at point B, would be 9.75, while the average cost of production for a larger firm producing 4 would only be 7.75. Thus, the economy would become less productively efficient, since the good is being produced at a higher average cost. In a situation with a downward-sloping average cost curve, two smaller firms will always have higher average costs of production than one larger firm for any quantity of total output. In addition, the antitrust authorities must worry that splitting the natural monopoly into pieces may be only the start of their problems. If one of the two firms

grows larger than the other, it will have lower average costs and may be able to drive its competitor out of the market. Alternatively, two firms in a market may discover subtle ways of coordinating their behavior and keeping prices high. Either way, the result will not be the greater competition that was desired.

A third alternative is that regulators may decide to set prices and quantities produced for this industry. The regulators will try to choose a point along the market demand curve that benefits both consumers and the broader social interest. Point C illustrates one tempting choice: the regulator requires that the firm produce the quantity of output where marginal cost crosses the demand curve at an output of 8, and charge the price of 3.5, which is equal to **marginal cost** at that point. This rule is appealing because it requires price to be set equal to marginal cost, which is what would occur in a perfectly competitive market, and it would assure consumers a higher quantity and lower price than at the monopoly choice A. In fact, efficient allocation of resources would occur at point C, since the value to the consumers of the last unit bought and sold in this market is equal to the marginal cost of producing it.

Attempting to bring about point C through force of regulation, however, runs into a severe difficulty. At point C, with an output of 8, a price of 3.5 is below the average cost of production, which is 5.7, and so if the firm charges a price of 3.5, it will be suffering losses. Unless the regulators or the government offer the firm an ongoing public subsidy (and there are numerous political problems with that option), the firm will lose money and go out of business.

Perhaps the most plausible option for the regulator is point F; that is, to set the price where AC crosses the demand curve at an output of 6 and a price of 6.5. This plan makes some sense at an intuitive level: let the natural monopoly charge enough to cover its average costs and earn a normal rate of profit, so that it can continue operating, but prevent the firm from raising prices and earning abnormally high monopoly profits, as it would at the monopoly choice A. Of course, determining this level of output and price with the political pressures, time constraints, and limited information of the real world is much harder than identifying the point on a graph. For more on the problems that can arise from a centrally determined price, see the discussion of price floors and price ceilings in [Demand and Supply](#).

## Cost-Plus versus Price Cap Regulation

Indeed, regulators of public utilities for many decades followed the general approach of attempting to choose a point like F in [\[link\]](#). They calculated the average cost of production for the water or electricity companies, added in an amount for the normal rate of profit the firm should expect to earn, and set the price for consumers accordingly. This method was known as **cost-plus regulation**.

Cost-plus regulation raises difficulties of its own. If producers are reimbursed for their costs, plus a bit more, then at a minimum, producers have less reason to be concerned with high costs—because they can just pass them along in higher prices. Worse, firms under cost-plus regulation even have an incentive to generate high costs by building huge factories or employing lots of staff, because what they can charge is linked to the costs they incur.

Thus, in the 1980s and 1990s, some regulators of public utilities began to use **price cap regulation**, where the regulator sets a price that the firm can charge over the next few years. A common pattern was to require a price that declined slightly over time. If the firm can find ways of reducing its costs more quickly than the price caps, it can make a high level of profits. However, if the firm cannot keep up with the price caps or suffers bad luck in the market, it may suffer losses. A few years down the road, the regulators will then set a new series of price caps based on the firm's performance.

Price cap regulation requires delicacy. It will not work if the price regulators set the price cap unrealistically low. It may not work if the market changes dramatically so that the firm is doomed to incurring losses no matter what it does—say, if energy prices rise dramatically on world markets, then the company selling natural gas or heating oil to homes may not be able to meet price caps that seemed reasonable a year or two ago. But if the regulators compare the prices with producers of the same good in other areas, they can, in effect, pressure a natural monopoly in one area to compete with the prices being charged in other areas. Moreover, the possibility of earning greater profits or experiencing losses—instead of having an average rate of profit locked in every year by cost-plus regulation—can provide the natural monopoly with incentives for efficiency and innovation.

With natural monopoly, market competition is unlikely to take root, so if consumers are not to suffer the high prices and restricted output of an unrestricted monopoly, government regulation will need to play a role. In attempting to design a system of price cap regulation with flexibility and incentive, government regulators do not have an easy task.

## Key Concepts and Summary

In the case of a natural monopoly, market competition will not work well and so, rather than allowing an unregulated monopoly to raise price and reduce output, the government may wish to regulate price and/or output. Common examples of regulation are public utilities, the regulated firms that often provide electricity and water service.

Cost-plus regulation refers to government regulation of a firm which sets the price that a firm can charge over a period of time by looking at the firm's accounting costs and then adding a normal rate of profit. Price cap regulation refers to government regulation of a firm where the government sets a price level several years in advance. In this case, the firm can either make high profits if it manages to produce at lower costs or sell a higher quantity than expected or suffer low profits or losses if costs are high or it sells less than expected.

## Self-Check Questions

### Exercise:

#### Problem:

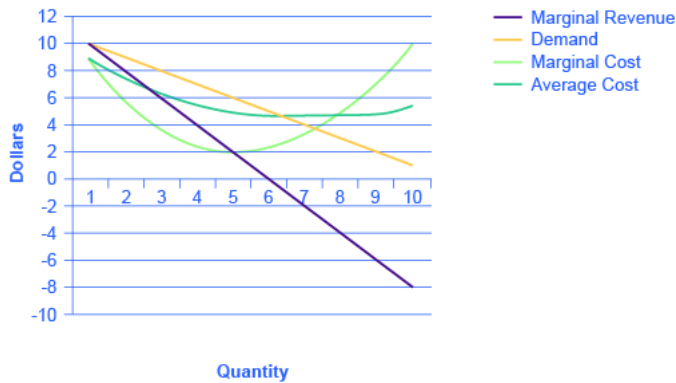
Urban transit systems, especially those with rail systems, typically experience significant economies of scale in operation. Consider the transit system whose data is given in the [\[link\]](#). Note that the quantity is in millions of riders.

Demand:	Quantity	1	2	3	4	5	6	7	8	9	10
	Price	10	9	8	7	6	5	4	3	2	1
	Marginal Revenue	10	8	6	4	2	0	-2	-4	-6	-8
Costs:	Marginal Cost	9	6	5	3	2	3	4	5	7	10
	Average Cost	9	7.5	6.7	5.8	5	4.7	4.6	4.6	4.9	5.4

Draw the demand, marginal revenue, marginal cost, and average cost curves. Do they have the normal shapes?

### Solution:

Yes, all curves have normal shapes.



### Exercise:

#### Problem:

From the graph you drew to answer [\[link\]](#), would you say this transit system is a natural monopoly? Justify.

#### Solution:

Yes it is a natural monopoly because average costs decline over the range that satisfies the market demand. For example, at the point where the demand curve and the average cost curve meet, there are economies of scale.

## Review Questions

### Exercise:

**Problem:** If public utilities are a natural monopoly, what would be the danger in deregulating them?

### Exercise:

#### Problem:

If public utilities are a natural monopoly, what would be the danger in splitting them up into a number of separate competing firms?

### Exercise:

**Problem:** What is cost-plus regulation?

### Exercise:

**Problem:** What is price cap regulation?

## Critical Thinking Questions

### Exercise:

#### Problem:

In the middle of the twentieth century, major U.S. cities had multiple competing city bus companies. Today, there is usually only one and it runs as a subsidized, regulated monopoly. What do you suppose caused the change?

### Exercise:

**Problem:**

Why are urban areas willing to subsidize urban transit systems? Does the argument for subsidies make sense to you?

**Problems**

Use [\[link\]](#) to answer the following questions.

**Exercise:****Problem:**

If the transit system was allowed to operate as an unregulated monopoly, what output would it supply and what price would it charge?

**Exercise:****Problem:**

If the transit system was regulated to operate with no subsidy (i.e., at zero economic profit), what approximate output would it supply and what approximate price would it charge?

**Exercise:****Problem:**

If the transit system was regulated to provide the most allocatively efficient quantity of output, what output would it supply and what price would it charge? What subsidy would be necessary to insure this efficient provision of transit services?

**Glossary**

cost-plus regulation

when regulators permit a regulated firm to cover its costs and to make a normal level of profit

price cap regulation

when the regulator sets a price that a firm cannot exceed over the next few years

## The Great Deregulation Experiment

By the end of this section, you will be able to:

- Evaluate the effectiveness of price regulation and antitrust policy
- Explain regulatory capture and its significance

Governments at all levels across the United States have regulated prices in a wide range of industries. In some cases, like water and electricity that have natural monopoly characteristics, there is some room in economic theory for such regulation. But once politicians are given a basis to intervene in markets and to choose prices and quantities, it is hard to know where to stop.

### Doubts about Regulation of Prices and Quantities

Beginning in the 1970s, it became clear to policymakers of all political leanings that the existing price regulation was not working well. The United States carried out a great policy experiment—the **deregulation** discussed in [Monopoly](#)—removing government controls over prices and quantities produced in airlines, railroads, trucking, intercity bus travel, natural gas, and bank interest rates. The Clear it Up discusses the outcome of deregulation in one industry in particular—airlines.

#### Note:

What are the results of airline deregulation?

Why did the pendulum swing in favor of deregulation? Consider the airline industry. In the early days of air travel, no airline could make a profit just by flying passengers. Airlines needed something else to carry and the Postal Service provided that something with airmail. And so the first U.S. government regulation of the airline industry happened through the Postal Service, when in 1926 the Postmaster General began giving airlines permission to fly certain routes based on the needs of mail delivery—and the airlines took some passengers along for the ride. In 1934, the Postmaster General was charged by the antitrust authorities with colluding

with the major airlines of that day to monopolize the nation's airways. In 1938, the Civil Aeronautics Board (CAB) was created to regulate airfares and routes instead. For 40 years, from 1938 to 1978, the CAB approved all fares, controlled all entry and exit, and specified which airlines could fly which routes. There was zero entry of new airlines on the main routes across the country for 40 years, because the CAB did not think it was necessary.

In 1978, the Airline Deregulation Act took the government out of the business of determining airfares and schedules. The new law shook up the industry. Famous old airlines like Pan American, Eastern, and Braniff went bankrupt and disappeared. Some new airlines like People Express were created—and then vanished.

The greater competition from deregulation reduced airfares by about one-third over the next two decades, saving consumers billions of dollars a year. The average flight used to take off with just half its seats full; now it is two-thirds full, which is far more efficient. Airlines have also developed hub-and-spoke systems, where planes all fly into a central hub city at a certain time and then depart. As a result, one can fly between any of the spoke cities with just one connection—and there is greater service to more cities than before deregulation. With lower fares and more service, the number of air passengers doubled from the late 1970s to the start of the 2000s—an increase that, in turn, doubled the number of jobs in the airline industry. Meanwhile, with the watchful oversight of government safety inspectors, commercial air travel has continued to get safer over time. The U.S. airline industry is far from perfect. For example, a string of mergers in recent years has raised concerns over how competition might be compromised.

One difficulty with government price regulation is what economists call **regulatory capture**, in which the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow. When the airline industry was being regulated, for example, it suggested appointees to the regulatory board, sent lobbyists to argue with the board, provided most of the information on which the board made decisions, and offered well-paid jobs to at least some of the people leaving the board. In this situation,

consumers can easily end up being not very well represented by the regulators. The result of regulatory capture is that government price regulation can often become a way for existing competitors to work together to reduce output, keep prices high, and limit competition.

## **The Effects of Deregulation**

Deregulation, both of airlines and of other industries, has its negatives. The greater pressure of competition led to entry and exit. When firms went bankrupt or contracted substantially in size, they laid off workers who had to find other jobs. Market competition is, after all, a full-contact sport.

A number of major accounting scandals involving prominent corporations such as Enron, Tyco International, and WorldCom led to the **Sarbanes-Oxley Act** in 2002. Sarbanes-Oxley was designed to increase confidence in financial information provided by public corporations to protect investors from accounting fraud.

The Great Recession which began in late 2007 and which the U.S. economy is still struggling to recover from was caused at least in part by a global financial crisis, which began in the United States. The key component of the crisis was the creation and subsequent failure of several types of unregulated financial assets, such as collateralized mortgage obligations (CMOs, a type of mortgage-backed security), and credit default swaps (CDSs, insurance contracts on assets like CMOs that provided a payoff even if the holder of the CDS did not own the CMO). Many of these assets were rated very safe by private credit rating agencies such as Standard & Poors, Moody's, and Fitch.

The collapse of the markets for these assets precipitated the financial crisis and led to the failure of Lehman Brothers, a major investment bank, numerous large commercial banks, such as Wachovia, and even the Federal National Mortgage Corporation (Fannie Mae), which had to be nationalized—that is, taken over by the federal government. One response to the financial crisis was the **Dodd-Frank Act**, which attempted major reforms of the financial system. The legislation's purpose, as noted on [dodd-frank.com](http://dodd-frank.com) is:

"To promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices. . . "

We will explore the financial crisis and the Great Recession in more detail in the macroeconomic chapters of this book, but for now it should be clear that many Americans have grown disenchanted with deregulation, at least of financial markets.

All market-based economies operate against a background of laws and regulations, including laws about enforcing contracts, collecting taxes, and protecting health and the environment. The government policies discussed in this chapter—like blocking certain anticompetitive mergers, ending restrictive practices, imposing price cap regulation on natural monopolies, and deregulation—demonstrate the role of government to strengthen the incentives that come with a greater degree of competition.

**Note:**

**More than Cooking, Heating, and Cooling**

What did the Federal Trade Commission (FTC) decide on the Kinder Morgan / El Paso Corporation merger? After careful examination, federal officials decided there was only one area of significant overlap that might provide the merged firm with strong market power. The FTC approved the merger, provided Kinder Morgan divest itself of the overlap area. Tallgrass purchased Kinder Morgan Interstate Gas Transmission, Trailblazer Pipeline Co. LLC, two processing facilities in Wyoming, and Kinder Morgan's 50 percent interest in the Rockies Express Pipeline to meet the FTC requirements. The FTC was attempting to strike a balance between potential cost reductions resulting from economies of scale and concentration of market power.

Did the price of natural gas decrease? Yes, rather significantly. In 2010, the wellhead price of natural gas was \$4.48 per thousand cubic foot; in 2012 the price had fallen to just \$2.66. Was the merger responsible for the large drop in price? The answer is uncertain. The larger contributor to the sharp drop in price was the overall increase in the supply of natural gas. More

and more natural gas was able to be recovered by fracturing shale deposits, a process called fracking. Fracking, which is controversial for environmental reasons, enabled the recovery of known reserves of natural gas that previously were not economically feasible to tap. Kinder Morgan's control of 80,000-plus miles of pipeline likely made moving the gas from wellheads to end users smoother and allowed for an even greater benefit from the increased supply.

## Key Concepts and Summary

The U.S. economy experienced a wave of deregulation in the late 1970s and early 1980s, when a number of government regulations that had set prices and quantities produced in a number of industries were eliminated. Major accounting scandals in the early 2000s and, more recently, the Great Recession have spurred new regulation to prevent similar occurrences in the future. Regulatory capture occurs when the industries being regulated end up having a strong influence over what regulations exist.

## Self-Check Questions

Use the following information to answer the next three questions. In the years before wireless phones, when telephone technology required having a wire running to every home, it seemed plausible that telephone service had diminishing average costs and might need to be regulated like a natural monopoly. For most of the twentieth century, the national U.S. phone company was AT&T, and the company functioned as a regulated monopoly. Think about the deregulation of the U.S. telecommunications industry that has happened over the last few decades. (This is not a research assignment, but a thought assignment based on what you have learned in this chapter.)

### Exercise:

**Problem:** What real world changes made the deregulation possible?

---

**Solution:**

Improvements in technology that allowed phone calls to be made via microwave transmission, communications satellites, and other wireless technologies.

**Exercise:**

**Problem:** What are some of the benefits of the deregulation?

---

**Solution:**

More consumer choice. Cheaper phone calls, especially long distance. Better-quality phone service in many cases. Cheaper, faster, and better-quality data transmission. Spin-off technologies like free Internet-based calling and video calling.

**Exercise:**

**Problem:** What might some of the negatives of deregulation be?

---

**Solution:**

More choice can sometimes make for difficult decisions—not knowing if you got the best plan for your situation, for example. Some phone service providers are less reliable than AT&T used to be.

## Review Questions

**Exercise:**

**Problem:**

What is deregulation? Name some industries that have been deregulated in the United States.

**Exercise:**

**Problem:** What is regulatory capture?

**Exercise:**

**Problem:**

Why does regulatory capture reduce the persuasiveness of the case for regulating industries for the benefit of consumers?

**Critical Thinking Questions****Exercise:****Problem:**

Deregulation, like all changes in government policy, always has pluses and minuses. What do you think some of the minuses might be for airline deregulation?

**Exercise:****Problem:**

Do you think it is possible for government to outlaw everything that businesses could do wrong? If so, why does government not do that? If not, how can regulation stay ahead of rogue businesses that push the limits of the system until it breaks?

**References**

Bishop, Todd. 2014. "Microsoft Exec Admits New Reality: Market Share No Longer 90% — It's 14%." *GeekWire*. Accessed March 27, 2015.  
<http://www.geekwire.com/2014/microsoft-exec-admits-new-reality-market-share-longer-90-14/>.

Catan, T., & DeZemmer, R. "Kinder-El Paso Merger to Face Antitrust Scrutiny," *Wall Street Journal*. October 19, 2011.

Collins, A. "Tallgrass Energy to Acquire Kinder Morgan Assets for \$1.8B." *The Middle Market*, Accessed August 2013.

<http://www.themiddlemarket.com/news/tallgrass-energy-to-acquire-kinder-morgan-assets-for-1-point-8-billion-232862-1.html>.

Conoco Phillips. "Why Natural Gas." Accessed August 2013.  
<http://www.powerincooperation.com/en/pages/useful.html?gclid=COXg7rH3hrGCFWlp7AodtkgA3g>.

De la Merced, M. (2012, August 20). "Kinder Morgan to Sell Assets to Tallgrass for \$1.8 Billion." *The New York Times*. August 20, 2012.

The Federal Trade Commission. n.d. "The Federal Trade Commission's (FTC) Mission." Accessed March 27, 2015.  
<https://www.ftc.gov/system/files/documents/reports/2014-one-page-ftc-performance-snapshot/150218fy14snapshot.pdf>.

Kahn, C. "Kinder Morgan to Buy El Paso Corp. for \$20.7B." *USA Today*. October 17, 2011.

Kinder Morgan. (2013). "Investor Information." Accessed August 2013.  
<http://www.kindermorgan.com/investor/>.

Rogowsky, Mark. 2014. "There'd Be No Wireless Wars Without The Blocked T-Mobile Merger, So Where Does That Leave Comcast-TWC?" *Forbes.com*. Accessed March 12, 2015.  
<http://www.forbes.com/sites/markrogowsky/2014/08/27/t-mobile-and-sprint-continue-to-battle-thanks-to-the-government/>.

U.S. Energy Information Administration (a). "Natural Gas Consumption by End User." Accessed May 31, 2013.  
[http://www.eia.gov/dnav/ng/ng\\_cons\\_sum\\_dcu\\_nus\\_a.htm](http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm).

U.S. Energy Information Administration (b). "Natural Gas Prices." Accessed June 28, 2013.  
[http://www.eia.gov/dnav/ng/ng\\_pri\\_sum\\_dcu\\_nus\\_a.htm](http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm).

The Wall Street Journal. 2015. "Auto Sales." Accessed April 10, 2015.  
[http://online.wsj.com/mdc/public/page/2\\_3022-autosales.html](http://online.wsj.com/mdc/public/page/2_3022-autosales.html).

## **Glossary**

### **regulatory capture**

when the firms supposedly being regulated end up playing a large role in setting the regulations that they will follow and as a result, they “capture” the people doing the regulation, usually through the promise of a job in that “regulated” industry once their term in government has ended.

## The Use of Mathematics in Principles of Economics

(This appendix should be consulted after first reading [Welcome to Economics!](#)) Economics is not math. There is no important concept in this course that cannot be explained without mathematics. That said, math is a tool that can be used to illustrate economic concepts. Remember the saying a picture is worth a thousand words? Instead of a picture, think of a graph. It is the same thing. Economists use models as the primary tool to derive insights about economic issues and problems. Math is one way of working with (or manipulating) economic models.

There are other ways of representing models, such as text or narrative. But why would you use your fist to bang a nail, if you had a hammer? Math has certain advantages over text. It disciplines your thinking by making you specify exactly what you mean. You can get away with fuzzy thinking in your head, but you cannot when you reduce a model to algebraic equations. At the same time, math also has disadvantages. Mathematical models are necessarily based on simplifying assumptions, so they are not likely to be perfectly realistic. Mathematical models also lack the nuances which can be found in narrative models. The point is that math is one tool, but it is not the only tool or even always the best tool economists can use. So what math will you need for this book? The answer is: little more than high school algebra and graphs. You will need to know:

- What a function is
- How to interpret the equation of a line (i.e., slope and intercept)
- How to manipulate a line (i.e., changing the slope or the intercept)
- How to compute and interpret a growth rate (i.e., percentage change)
- How to read and manipulate a graph

In this text, we will use the easiest math possible, and we will introduce it in this appendix. So if you find some math in the book that you cannot follow, come back to this appendix to review. Like most things, math has diminishing returns. A little math ability goes a long way; the more advanced math you bring in, the less additional knowledge that will get you. That said, if you are going to major in economics, you should consider learning a little calculus. It will be worth your while in terms of helping you learn advanced economics more quickly.

## Algebraic Models

Often economic models (or parts of models) are expressed in terms of mathematical functions. What is a function? A function describes a relationship. Sometimes the relationship is a definition. For example (using words), your professor is Adam Smith. This could be expressed as Professor = Adam Smith. Or Friends = Bob + Shawn + Margaret.

Often in economics, functions describe cause and effect. The variable on the left-hand side is what is being explained (“the effect”). On the right-hand side is what is doing the explaining (“the causes”). For example, suppose your GPA was determined as follows:

**Equation:**

$$\text{GPA} = 0.25 \times \text{combined\_SAT} + 0.25 \times \text{class\_attendance} + 0.50 \times \text{hours\_spent\_studying}$$

This equation states that your GPA depends on three things: your combined SAT score, your class attendance, and the number of hours you spend studying. It also says that study time is twice as

important (0.50) as either combined\_SAT score (0.25) or class\_attendance (0.25). If this relationship is true, how could you raise your GPA? By not skipping class and studying more. Note that you cannot do anything about your SAT score, since if you are in college, you have (presumably) already taken the SATs.

Of course, economic models express relationships using economic variables, like  $\text{Budget} = \text{money\_spent\_on\_econ\_books} + \text{money\_spent\_on\_music}$ , assuming that the only things you buy are economics books and music.

Most of the relationships we use in this course are expressed as linear equations of the form:

**Equation:**

$$y = b + mx$$

**Expressing Equations Graphically**

Graphs are useful for two purposes. The first is to express equations visually, and the second is to display statistics or data. This section will discuss expressing equations visually.

To a mathematician or an economist, a variable is the name given to a quantity that may assume a range of values. In the equation of a line presented above, x and y are the variables, with x on the horizontal axis and y on the vertical axis, and b and m representing factors that determine the shape of the line. To see how this equation works, consider a numerical example:

**Equation:**

$$y = 9 + 3x$$

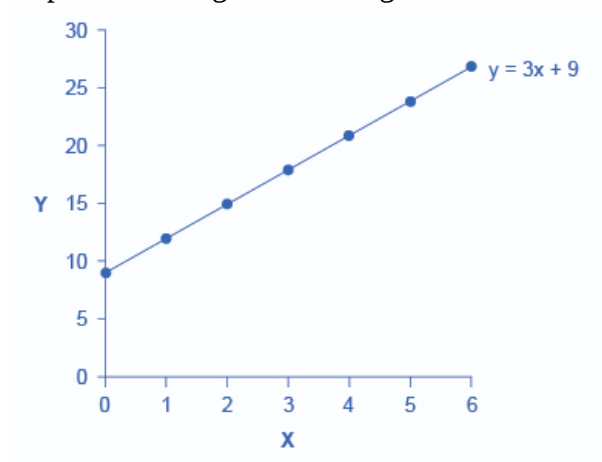
In this equation for a specific line, the b term has been set equal to 9 and the m term has been set equal to 3. [\[link\]](#) shows the values of x and y for this given equation. [\[link\]](#) shows this equation, and these values, in a graph. To construct the table, just plug in a series of different values for x, and then calculate what value of y results. In the figure, these points are plotted and a line is drawn through them.

x	y
0	9
1	12
2	15
3	18

x	y
4	21
5	24
6	27

Values for the Slope Intercept Equation

Slope and the Algebra of Straight Lines



This line graph has  $x$  on the horizontal axis and  $y$  on the vertical axis. The  $y$ -intercept—that is, the point where the line intersects the  $y$ -axis—is 9. The slope of the line is 3; that is, there is a rise of 3 on the vertical axis for every increase of 1 on the horizontal axis. The slope is the same all along a straight line.

This example illustrates how the  $b$  and  $m$  terms in an equation for a straight line determine the shape of the line. The  $b$  term is called the  $y$ -intercept. The reason for this name is that, if  $x = 0$ , then the  $b$  term will reveal where the line intercepts, or crosses, the  $y$ -axis. In this example, the line hits the vertical axis at 9. The  $m$  term in the equation for the line is the slope. Remember that slope is defined as rise over run; more specifically, the slope of a line from one point to another is the change in the vertical axis divided by the change in the horizontal axis. In this example, each time the  $x$  term increases by one (the run), the  $y$  term rises by three. Thus, the slope of this line is three. Specifying a  $y$ -intercept and a slope—that is, specifying  $b$  and  $m$  in the equation for a line—will identify a specific line. Although it is rare for real-world data points to arrange themselves as an exact straight line, it often turns out that a straight line can offer a reasonable approximation of actual data.

## Interpreting the Slope

The concept of slope is very useful in economics, because it measures the relationship between two variables. A positive slope means that two variables are positively related; that is, when  $x$  increases, so does  $y$ , or when  $x$  decreases,  $y$  decreases also. Graphically, a positive slope means that as a line on the line graph moves from left to right, the line rises. The length-weight relationship, shown in [\[link\]](#) later in this Appendix, has a positive slope. We will learn in other chapters that price and quantity supplied have a positive relationship; that is, firms will supply more when the price is higher.

A negative slope means that two variables are negatively related; that is, when  $x$  increases,  $y$  decreases, or when  $x$  decreases,  $y$  increases. Graphically, a negative slope means that, as the line on the line graph moves from left to right, the line falls. The altitude-air density relationship, shown in [\[link\]](#) later in this appendix, has a negative slope. We will learn that price and quantity demanded have a negative relationship; that is, consumers will purchase less when the price is higher.

A slope of zero means that there is no relationship between  $x$  and  $y$ . Graphically, the line is flat; that is, zero rise over the run. [\[link\]](#) of the unemployment rate, shown later in this appendix, illustrates a common pattern of many line graphs: some segments where the slope is positive, other segments where the slope is negative, and still other segments where the slope is close to zero.

The slope of a straight line between two points can be calculated in numerical terms. To calculate slope, begin by designating one point as the “starting point” and the other point as the “end point” and then calculating the rise over run between these two points. As an example, consider the slope of the air density graph between the points representing an altitude of 4,000 meters and an altitude of 6,000 meters:

Rise: Change in variable on vertical axis (end point minus original point)

**Equation:**

$$\begin{aligned} &= 0.100 - 0.307 \\ &= -0.207 \end{aligned}$$

Run: Change in variable on horizontal axis (end point minus original point)

**Equation:**

$$\begin{aligned} &= 6,000 - 4,000 \\ &= 2,000 \end{aligned}$$

Thus, the slope of a straight line between these two points would be that from the altitude of 4,000 meters up to 6,000 meters, the density of the air decreases by approximately 0.1 kilograms/cubic meter for each of the next 1,000 meters

Suppose the slope of a line were to increase. Graphically, that means it would get steeper. Suppose the slope of a line were to decrease. Then it would get flatter. These conditions are true whether or not the slope was positive or negative to begin with. A higher positive slope means a steeper upward tilt to the line, while a smaller positive slope means a flatter upward tilt to the line. A

negative slope that is larger in absolute value (that is, more negative) means a steeper downward tilt to the line. A slope of zero is a horizontal flat line. A vertical line has an infinite slope.

Suppose a line has a larger intercept. Graphically, that means it would shift out (or up) from the old origin, parallel to the old line. If a line has a smaller intercept, it would shift in (or down), parallel to the old line.

### **Solving Models with Algebra**

Economists often use models to answer a specific question, like: What will the unemployment rate be if the economy grows at 3% per year? Answering specific questions requires solving the “system” of equations that represent the model.

Suppose the demand for personal pizzas is given by the following equation:

**Equation:**

$$Q_d = 16 - 2P$$

where  $Q_d$  is the amount of personal pizzas consumers want to buy (i.e., quantity demanded), and  $P$  is the price of pizzas. Suppose the supply of personal pizzas is:

**Equation:**

$$Q_s = 2 + 5P$$

where  $Q_s$  is the amount of pizza producers will supply (i.e., quantity supplied).

Finally, suppose that the personal pizza market operates where supply equals demand, or

**Equation:**

$$Q_d = Q_s$$

We now have a system of three equations and three unknowns ( $Q_d$ ,  $Q_s$ , and  $P$ ), which we can solve with algebra:

Since  $Q_d = Q_s$ , we can set the demand and supply equation equal to each other:

**Equation:**

$$\begin{aligned} Q_d &= Q_s \\ 16 - 2P &= 2 + 5P \end{aligned}$$

Subtracting 2 from both sides and adding  $2P$  to both sides yields:

**Equation:**

$$\begin{aligned}
 16 - 2P - 2 &= 2 + 5P - 2 \\
 14 - 2P &= 5P \\
 14 - 2P + 2P &= 5P + 2P \\
 14 &= 7P \\
 \frac{14}{7} &= \frac{7P}{7} \\
 2 &= P
 \end{aligned}$$

In other words, the price of each personal pizza will be \$2. How much will consumers buy?

Taking the price of \$2, and plugging it into the demand equation, we get:

**Equation:**

$$\begin{aligned}
 Q_d &= 16 - 2P \\
 &= 16 - 2(2) \\
 &= 16 - 4 \\
 &= 12
 \end{aligned}$$

So if the price is \$2 each, consumers will purchase 12. How much will producers supply? Taking the price of \$2, and plugging it into the supply equation, we get:

**Equation:**

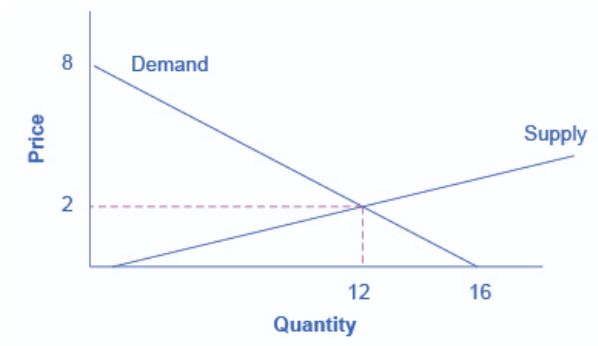
$$\begin{aligned}
 Q_s &= 2 + 5P \\
 &= 2 + 5(2) \\
 &= 2 + 10 \\
 &= 12
 \end{aligned}$$

So if the price is \$2 each, producers will supply 12 personal pizzas. This means we did our math correctly, since  $Q_d = Q_s$ .

### Solving Models with Graphs

If algebra is not your forte, you can get the same answer by using graphs. Take the equations for  $Q_d$  and  $Q_s$  and graph them on the same set of axes as shown in [\[link\]](#). Since  $P$  is on the vertical axis, it is easiest if you solve each equation for  $P$ . The demand curve is then  $P = 8 - 0.5Q_d$  and the supply curve is  $P = -0.4 + 0.2Q_s$ . Note that the vertical intercepts are 8 and  $-0.4$ , and the slopes are  $-0.5$  for demand and  $0.2$  for supply. If you draw the graphs carefully, you will see that where they cross ( $Q_s = Q_d$ ), the price is \$2 and the quantity is 12, just like the algebra predicted.

Supply and Demand Graph



The equations for  $Q_d$  and  $Q_s$  are displayed graphically by the sloped lines.

We will use graphs more frequently in this book than algebra, but now you know the math behind the graphs.

## Growth Rates

Growth rates are frequently encountered in real world economics. A growth rate is simply the percentage change in some quantity. It could be your income. It could be a business's sales. It could be a nation's GDP. The formula for computing a growth rate is straightforward:

**Equation:**

$$\text{Percentage change} = \frac{\text{Change in quantity}}{\text{Quantity}}$$

Suppose your job pays \$10 per hour. Your boss, however, is so impressed with your work that he gives you a \$2 per hour raise. The percentage change (or growth rate) in your pay is  $\$2/\$10 = 0.20$  or 20%.

To compute the growth rate for data over an extended period of time, for example, the average annual growth in GDP over a decade or more, the denominator is commonly defined a little differently. In the previous example, we defined the quantity as the initial quantity—or the quantity when we started. This is fine for a one-time calculation, but when we compute the growth over and over, it makes more sense to define the quantity as the average quantity over the period in question, which is defined as the quantity halfway between the initial quantity and the next quantity. This is harder to explain in words than to show with an example. Suppose a nation's GDP was \$1 trillion in 2005 and \$1.03 trillion in 2006. The growth rate between 2005 and 2006 would be the change in GDP ( $\$1.03 \text{ trillion} - \$1.00 \text{ trillion}$ ) divided by the average GDP between 2005 and 2006 ( $\$1.03 \text{ trillion} + \$1.00 \text{ trillion} / 2$ ). In other words:

**Equation:**

$$\begin{aligned}
&= \frac{\$1.03 \text{ trillion} - \$1.00 \text{ trillion}}{(\$1.03 \text{ trillion} + \$1.00 \text{ trillion}) / 2} \\
&= \frac{0.03}{1.015} \\
&= 0.0296 \\
&= 2.96\% \text{ growth}
\end{aligned}$$

Note that if we used the first method, the calculation would be  $(\$1.03 \text{ trillion} - \$1.00 \text{ trillion}) / \$1.00 \text{ trillion} = 3\%$  growth, which is approximately the same as the second, more complicated method. If you need a rough approximation, use the first method. If you need accuracy, use the second method.

A few things to remember: A positive growth rate means the quantity is growing. A smaller growth rate means the quantity is growing more slowly. A larger growth rate means the quantity is growing more quickly. A negative growth rate means the quantity is decreasing.

The same change over times yields a smaller growth rate. If you got a \$2 raise each year, in the first year the growth rate would be  $\$2/\$10 = 20\%$ , as shown above. But in the second year, the growth rate would be  $\$2/\$12 = 0.167$  or 16.7% growth. In the third year, the same \$2 raise would correspond to a  $\$2/\$14 = 14.2\%$ . The moral of the story is this: To keep the growth rate the same, the change must increase each period.

## Displaying Data Graphically and Interpreting the Graph

Graphs are also used to display data or evidence. Graphs are a method of presenting numerical patterns. They condense detailed numerical information into a visual form in which relationships and numerical patterns can be seen more easily. For example, which countries have larger or smaller populations? A careful reader could examine a long list of numbers representing the populations of many countries, but with over 200 nations in the world, searching through such a list would take concentration and time. Putting these same numbers on a graph can quickly reveal population patterns. Economists use graphs both for a compact and readable presentation of groups of numbers and for building an intuitive grasp of relationships and connections.

Three types of graphs are used in this book: line graphs, pie graphs, and bar graphs. Each is discussed below. We also provide warnings about how graphs can be manipulated to alter viewers' perceptions of the relationships in the data.

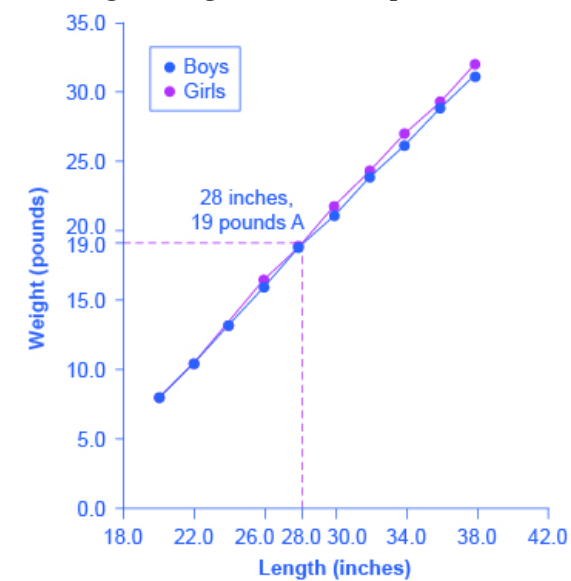
### Line Graphs

The graphs we have discussed so far are called line graphs, because they show a relationship between two variables: one measured on the horizontal axis and the other measured on the vertical axis.

Sometimes it is useful to show more than one set of data on the same axes. The data in [\[link\]](#) is displayed in [\[link\]](#) which shows the relationship between two variables: length and median weight for American baby boys and girls during the first three years of life. (The median means that half of all babies weigh more than this and half weigh less.) The line graph measures length in inches on the horizontal axis and weight in pounds on the vertical axis. For example, point A on the figure shows that a boy who is 28 inches long will have a median weight of about 19 pounds. One line on

the graph shows the length-weight relationship for boys and the other line shows the relationship for girls. This kind of graph is widely used by healthcare providers to check whether a child’s physical development is roughly on track.

The Length-Weight Relationship for American Boys and Girls



The line graph shows the relationship between height and weight for boys and girls from birth to 3 years. Point A, for example, shows that a boy of 28 inches in height (measured on the horizontal axis) is typically 19 pounds in weight (measured on the vertical axis). These data apply only to children in the first three years of life.

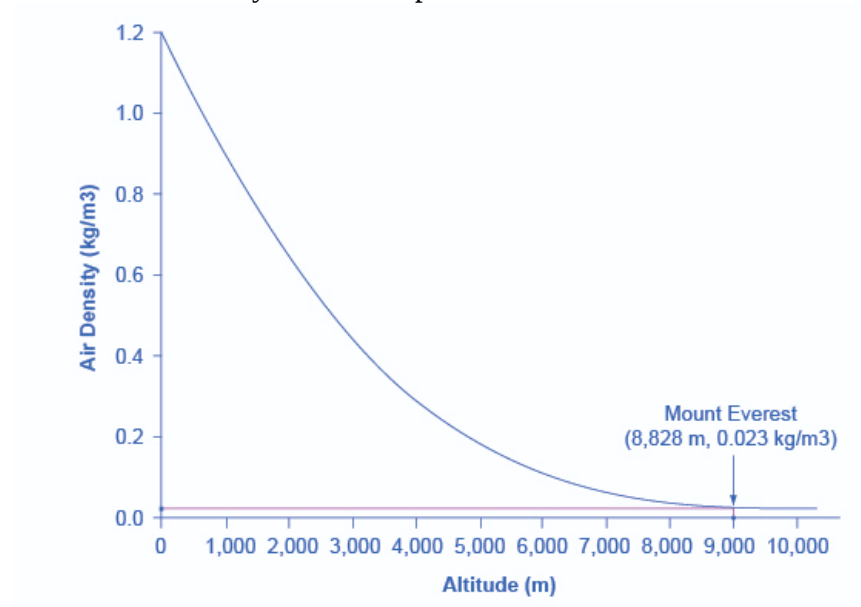
Boys from Birth to 36 Months		Girls from Birth to 36 Months	
Length (inches)	Weight (pounds)	Length (inches)	Weight (pounds)
20.0	8.0	20.0	7.9
22.0	10.5	22.0	10.5
24.0	13.5	24.0	13.2

Boys from Birth to 36 Months		Girls from Birth to 36 Months	
26.0	16.4	26.0	16.0
28.0	19.0	28.0	18.8
30.0	21.8	30.0	21.2
32.0	24.3	32.0	24.0
34.0	27.0	34.0	26.2
36.0	29.3	36.0	28.9
38.0	32.0	38.0	31.3

### Length to Weight Relationship for American Boys and Girls

Not all relationships in economics are linear. Sometimes they are curves. [\[link\]](#) presents another example of a line graph, representing the data from [\[link\]](#). In this case, the line graph shows how thin the air becomes when you climb a mountain. The horizontal axis of the figure shows altitude, measured in meters above sea level. The vertical axis measures the density of the air at each altitude. Air density is measured by the weight of the air in a cubic meter of space (that is, a box measuring one meter in height, width, and depth). As the graph shows, air pressure is heaviest at ground level and becomes lighter as you climb. [\[link\]](#) shows that a cubic meter of air at an altitude of 500 meters weighs approximately one kilogram (about 2.2 pounds). However, as the altitude increases, air density decreases. A cubic meter of air at the top of Mount Everest, at about 8,828 meters, would weigh only 0.023 kilograms. The thin air at high altitudes explains why many mountain climbers need to use oxygen tanks as they reach the top of a mountain.

### Altitude-Air Density Relationship



This line graph shows the relationship between altitude, measured in meters above sea level, and air density, measured in kilograms of air per cubic meter. As altitude rises, air density declines. The point at the top of Mount Everest has an altitude of approximately 8,828 meters above sea level (the horizontal axis) and air density of 0.023 kilograms per cubic meter (the vertical axis).

Altitude (meters)	Air Density (kg/cubic meters)
0	1.200
500	1.093
1,000	0.831
1,500	0.678
2,000	0.569
2,500	0.484
3,000	0.415
3,500	0.357
4,000	0.307
4,500	0.231
5,000	0.182
5,500	0.142
6,000	0.100
6,500	0.085
7,000	0.066
7,500	0.051

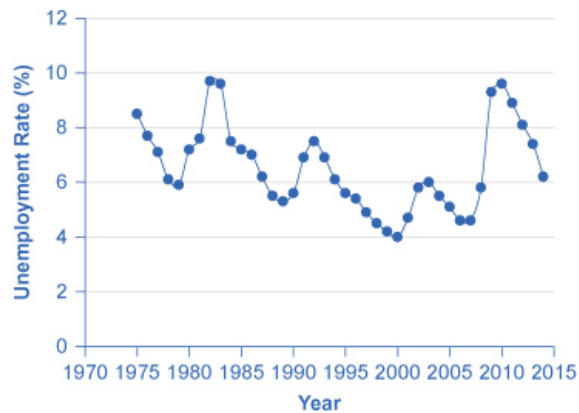
Altitude (meters)	Air Density (kg/cubic meters)
8,000	0.041
8,500	0.025
9,000	0.022
9,500	0.019
10,000	0.014

### Altitude to Air Density Relationship

The length-weight relationship and the altitude-air density relationships in these two figures represent averages. If you were to collect actual data on air pressure at different altitudes, the same altitude in different geographic locations will have slightly different air density, depending on factors like how far you are from the equator, local weather conditions, and the humidity in the air. Similarly, in measuring the height and weight of children for the previous line graph, children of a particular height would have a range of different weights, some above average and some below. In the real world, this sort of variation in data is common. The task of a researcher is to organize that data in a way that helps to understand typical patterns. The study of statistics, especially when combined with computer statistics and spreadsheet programs, is a great help in organizing this kind of data, plotting line graphs, and looking for typical underlying relationships. For most economics and social science majors, a statistics course will be required at some point.

One common line graph is called a time series, in which the horizontal axis shows time and the vertical axis displays another variable. Thus, a time series graph shows how a variable changes over time. [\[link\]](#) shows the unemployment rate in the United States since 1975, where unemployment is defined as the percentage of adults who want jobs and are looking for a job, but cannot find one. The points for the unemployment rate in each year are plotted on the graph, and a line then connects the points, showing how the unemployment rate has moved up and down since 1975. The line graph makes it easy to see, for example, that the highest unemployment rate during this time period was slightly less than 10% in the early 1980s and 2010, while the unemployment rate declined from the early 1990s to the end of the 1990s, before rising and then falling back in the early 2000s, and then rising sharply during the recession from 2008–2009.

U.S. Unemployment Rate, 1975–2014



This graph provides a quick visual summary of unemployment data. With a graph like this, it is easy to spot the times of high unemployment and of low unemployment.

## Pie Graphs

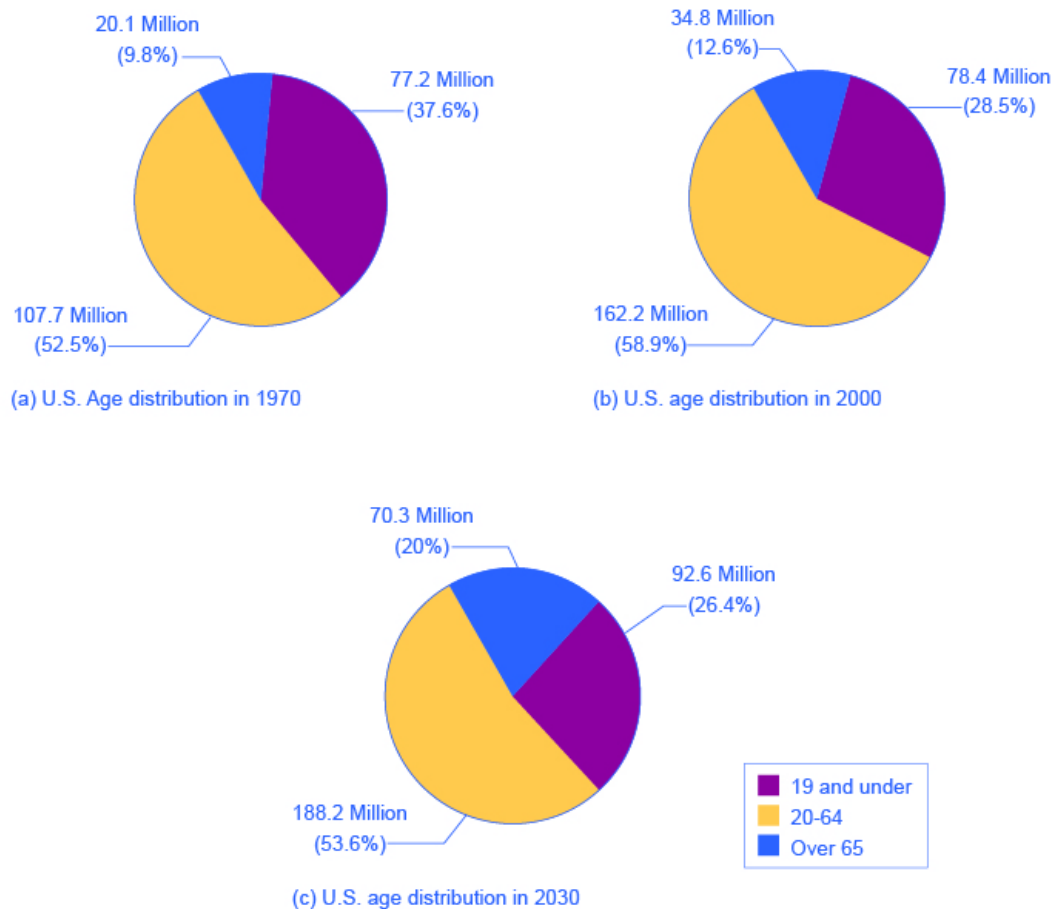
A pie graph (sometimes called a pie chart) is used to show how an overall total is divided into parts. A circle represents a group as a whole. The slices of this circular “pie” show the relative sizes of subgroups.

[\[link\]](#) shows how the U.S. population was divided among children, working age adults, and the elderly in 1970, 2000, and what is projected for 2030. The information is first conveyed with numbers in [\[link\]](#), and then in three pie charts. The first column of [\[link\]](#) shows the total U.S. population for each of the three years. Columns 2–4 categorize the total in terms of age groups—from birth to 18 years, from 19 to 64 years, and 65 years and above. In columns 2–4, the first number shows the actual number of people in each age category, while the number in parentheses shows the percentage of the total population comprised by that age group.

Year	Total Population	19 and Under	20–64 years	Over 65
1970	205.0 million	77.2 (37.6%)	107.7 (52.5%)	20.1 (9.8%)
2000	275.4 million	78.4 (28.5%)	162.2 (58.9%)	34.8 (12.6%)
2030	351.1 million	92.6 (26.4%)	188.2 (53.6%)	70.3 (20.0%)

U.S. Age Distribution, 1970, 2000, and 2030 (projected)

## Pie Graphs of the U.S. Age Distribution (numbers in millions)



The three pie graphs illustrate the division of total population into three age groups for the three different years.

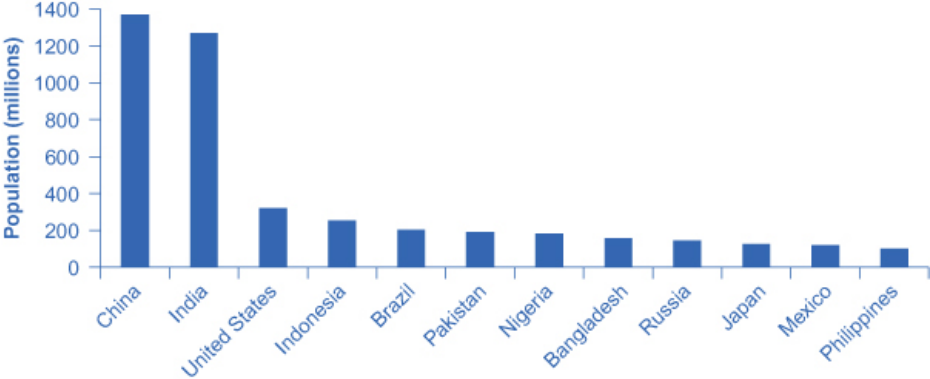
In a pie graph, each slice of the pie represents a share of the total, or a percentage. For example, 50% would be half of the pie and 20% would be one-fifth of the pie. The three pie graphs in [\[link\]](#) show that the share of the U.S. population 65 and over is growing. The pie graphs allow you to get a feel for the relative size of the different age groups from 1970 to 2000 to 2030, without requiring you to slog through the specific numbers and percentages in the table. Some common examples of how pie graphs are used include dividing the population into groups by age, income level, ethnicity, religion, occupation; dividing different firms into categories by size, industry, number of employees; and dividing up government spending or taxes into its main categories.

## Bar Graphs

A bar graph uses the height of different bars to compare quantities. [\[link\]](#) lists the 12 most populous countries in the world. [\[link\]](#) provides this same data in a bar graph. The height of the bars corresponds to the population of each country. Although you may know that China and India

are the most populous countries in the world, seeing how the bars on the graph tower over the other countries helps illustrate the magnitude of the difference between the sizes of national populations.

Leading Countries of the World by Population, 2015 (in millions)



The graph shows the 12 countries of the world with the largest populations. The height of the bars in the bar graph shows the size of the population for each country.

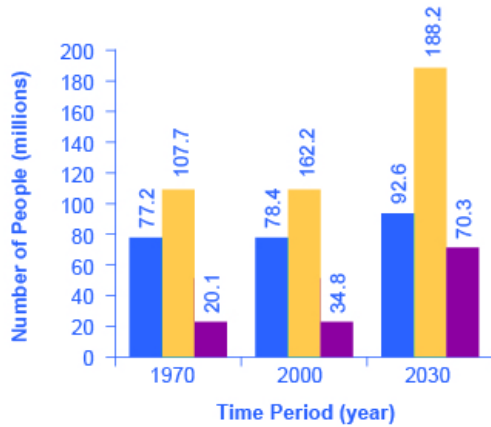
Country	Population
China	1,369
India	1,270
United States	321
Indonesia	255
Brazil	204
Pakistan	190
Nigeria	184
Bangladesh	158
Russia	146
Japan	127

Country	Population
Mexico	121
Philippines	101

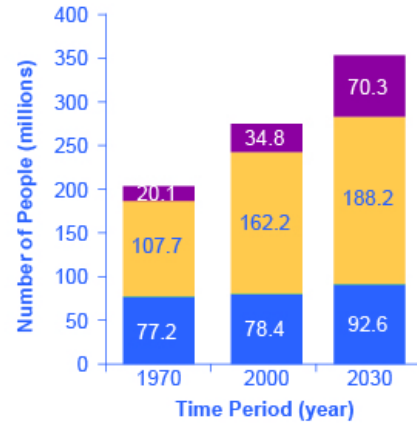
### Leading 12 Countries of the World by Population

Bar graphs can be subdivided in a way that reveals information similar to that we can get from pie charts. [\[link\]](#) offers three bar graphs based on the information from [\[link\]](#) about the U.S. age distribution in 1970, 2000, and 2030. [\[link\]](#) (a) shows three bars for each year, representing the total number of persons in each age bracket for each year. [\[link\]](#) (b) shows just one bar for each year, but the different age groups are now shaded inside the bar. In [\[link\]](#) (c), still based on the same data, the vertical axis measures percentages rather than the number of persons. In this case, all three bar graphs are the same height, representing 100% of the population, with each bar divided according to the percentage of population in each age group. It is sometimes easier for a reader to run his or her eyes across several bar graphs, comparing the shaded areas, rather than trying to compare several pie graphs.

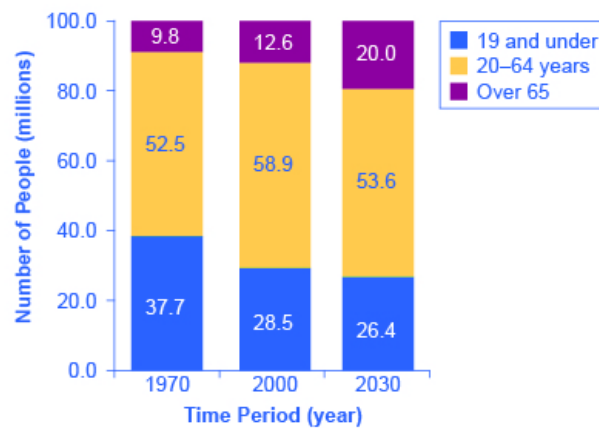
### U.S. Population with Bar Graphs



(a) Bars for separate age groups



(b) Bars show total population divided into age groups



(c) Bars show total population divided into percentages

Population data can be represented in different ways. (a) Shows three bars for each year, representing the total number of persons in each age bracket for each year. (b) Shows just one bar for each year, but the different age groups are now shaded inside the bar. (c) Sets the vertical axis as a measure of percentages rather than the number of persons. All three bar graphs are the same height and each bar is divided according to the percentage of population in each age group.

[\[link\]](#) and [\[link\]](#) show how the bars can represent countries or years, and how the vertical axis can represent a numerical or a percentage value. Bar graphs can also compare size, quantity, rates, distances, and other quantitative categories.

## Comparing Line Graphs with Pie Charts and Bar Graphs

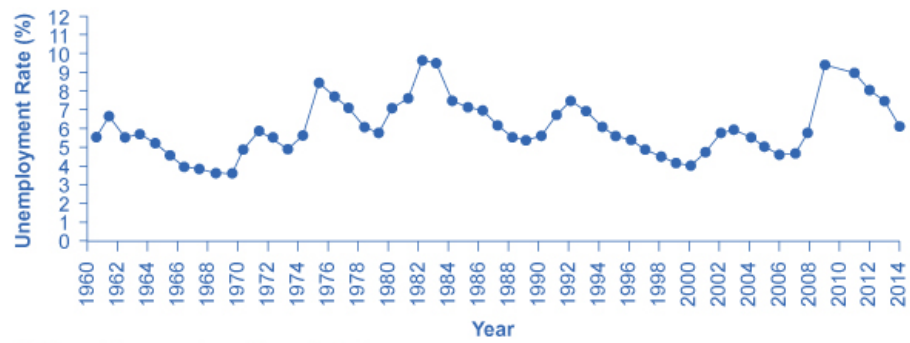
Now that you are familiar with pie graphs, bar graphs, and line graphs, how do you know which graph to use for your data? Pie graphs are often better than line graphs at showing how an overall group is divided. However, if a pie graph has too many slices, it can become difficult to interpret.

Bar graphs are especially useful when comparing quantities. For example, if you are studying the populations of different countries, as in [\[link\]](#), bar graphs can show the relationships between the population sizes of multiple countries. Not only can it show these relationships, but it can also show breakdowns of different groups within the population.

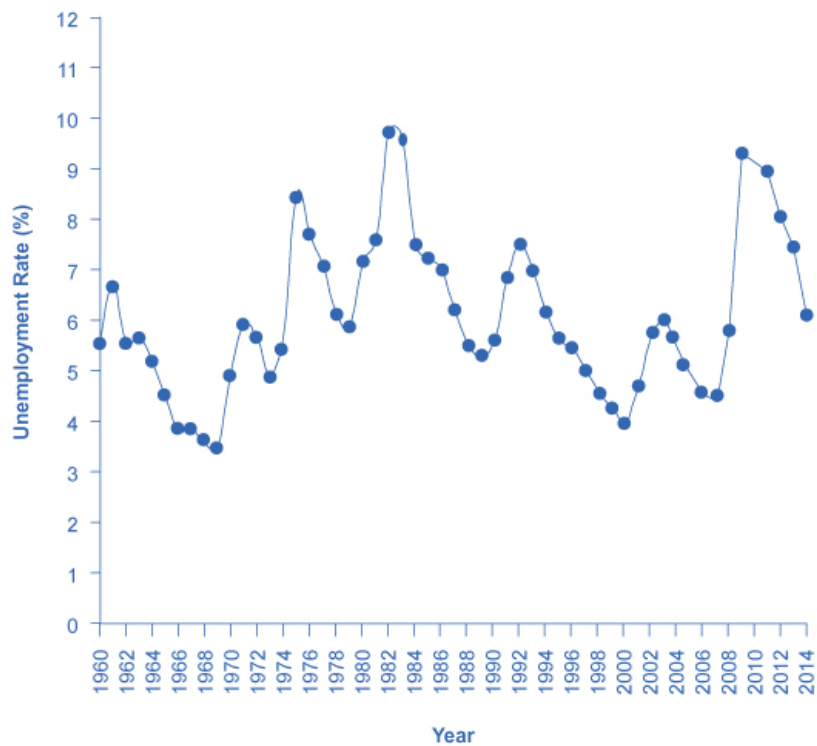
A line graph is often the most effective format for illustrating a relationship between two variables that are both changing. For example, time series graphs can show patterns as time changes, like the unemployment rate over time. Line graphs are widely used in economics to present continuous data about prices, wages, quantities bought and sold, the size of the economy.

### **How Graphs Can Be Misleading**

Graphs not only reveal patterns; they can also alter how patterns are perceived. To see some of the ways this can be done, consider the line graphs of [\[link\]](#), [\[link\]](#), and [\[link\]](#). These graphs all illustrate the unemployment rate—but from different perspectives.

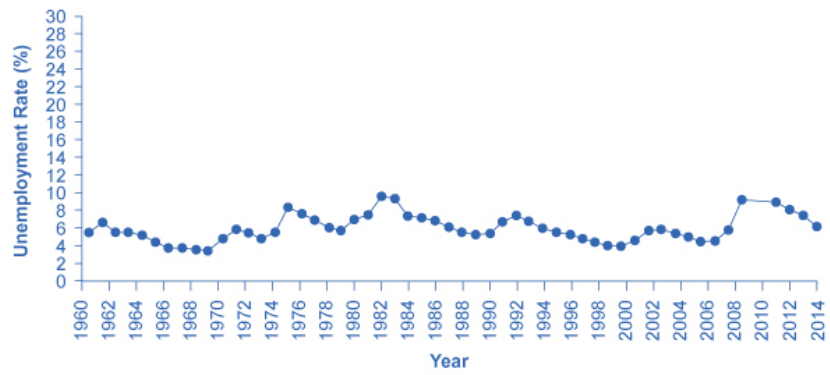


(a) Unemployment rate, wide and short

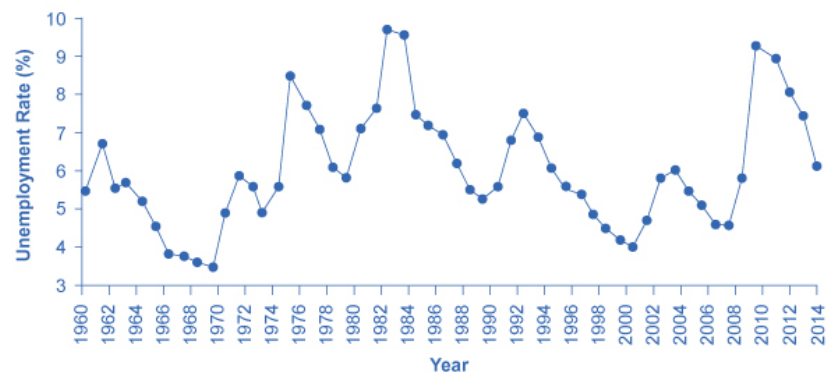


(b) Unemployment rate, narrow and tall

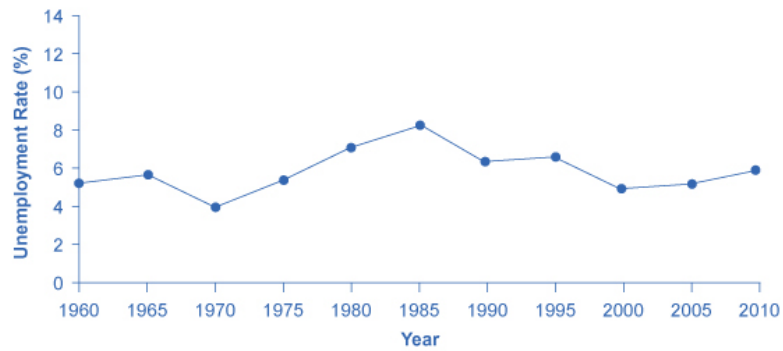
Presenting Unemployment Rates in Different Ways, All of Them Accurate



(c) Unemployment rate, with wider range of numbers on vertical axis



(d) Unemployment rate, with smaller range of numbers on vertical axis



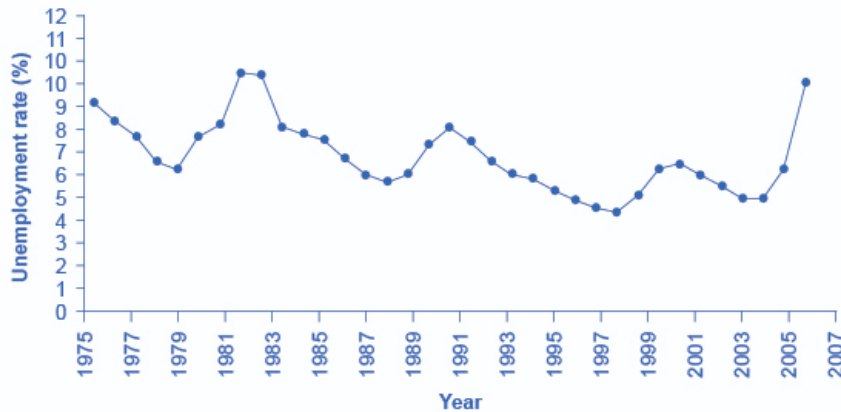
(e) Unemployment rate, five-year averages

Simply changing the width and height of the area in which data is displayed can alter the perception of the data.

Presenting Unemployment Rates in Different Ways, All of Them Accurate



(f) Unemployment rate, monthly data



(g) Unemployment rates, since 1975 only

Simply changing the width and height of the area in which data is displayed can alter the perception of the data.

Suppose you wanted a graph which gives the impression that the rise in unemployment in 2009 was not all that large, or all that extraordinary by historical standards. You might choose to present your data as in [\[link\]](#) (a). [\[link\]](#) (a) includes much of the same data presented earlier in [\[link\]](#), but stretches the horizontal axis out longer relative to the vertical axis. By spreading the graph wide and flat, the visual appearance is that the rise in unemployment is not so large, and is similar to some past rises in unemployment. Now imagine you wanted to emphasize how unemployment spiked substantially higher in 2009. In this case, using the same data, you can stretch the vertical axis out relative to the horizontal axis, as in [\[link\]](#) (b), which makes all rises and falls in unemployment appear larger.

A similar effect can be accomplished without changing the length of the axes, but by changing the scale on the vertical axis. In [\[link\]](#) (c), the scale on the vertical axis runs from 0% to 30%, while in [\[link\]](#) (d), the vertical axis runs from 3% to 10%. Compared to [\[link\]](#), where the vertical scale runs

from 0% to 12%, [\[link\]](#) (c) makes the fluctuation in unemployment look smaller, while [\[link\]](#) (d) makes it look larger.

Another way to alter the perception of the graph is to reduce the amount of variation by changing the number of points plotted on the graph. [\[link\]](#) (e) shows the unemployment rate according to five-year averages. By averaging out some of the year- to-year changes, the line appears smoother and with fewer highs and lows. In reality, the unemployment rate is reported monthly, and [\[link\]](#) (f) shows the monthly figures since 1960, which fluctuate more than the five-year average. [\[link\]](#) (f) is also a vivid illustration of how graphs can compress lots of data. The graph includes monthly data since 1960, which over almost 50 years, works out to nearly 600 data points. Reading that list of 600 data points in numerical form would be hypnotic. You can, however, get a good intuitive sense of these 600 data points very quickly from the graph.

A final trick in manipulating the perception of graphical information is that, by choosing the starting and ending points carefully, you can influence the perception of whether the variable is rising or falling. The original data show a general pattern with unemployment low in the 1960s, but spiking up in the mid-1970s, early 1980s, early 1990s, early 2000s, and late 2000s. [\[link\]](#) (g), however, shows a graph that goes back only to 1975, which gives an impression that unemployment was more-or-less gradually falling over time until the 2009 recession pushed it back up to its “original” level—which is a plausible interpretation if one starts at the high point around 1975.

These kinds of tricks—or shall we just call them “presentation choices”—are not limited to line graphs. In a pie chart with many small slices and one large slice, someone must decide what categories should be used to produce these slices in the first place, thus making some slices appear bigger than others. If you are making a bar graph, you can make the vertical axis either taller or shorter, which will tend to make variations in the height of the bars appear more or less.

Being able to read graphs is an essential skill, both in economics and in life. A graph is just one perspective or point of view, shaped by choices such as those discussed in this section. Do not always believe the first quick impression from a graph. View with caution.

## Key Concepts and Summary

Math is a tool for understanding economics and economic relationships can be expressed mathematically using algebra or graphs. The algebraic equation for a line is  $y = b + mx$ , where  $x$  is the variable on the horizontal axis and  $y$  is the variable on the vertical axis, the  $b$  term is the  $y$ -intercept and the  $m$  term is the slope. The slope of a line is the same at any point on the line and it indicates the relationship (positive, negative, or zero) between two economic variables.

Economic models can be solved algebraically or graphically. Graphs allow you to illustrate data visually. They can illustrate patterns, comparisons, trends, and apportionment by condensing the numerical data and providing an intuitive sense of relationships in the data. A line graph shows the relationship between two variables: one is shown on the horizontal axis and one on the vertical axis. A pie graph shows how something is allotted, such as a sum of money or a group of people. The size of each slice of the pie is drawn to represent the corresponding percentage of the whole. A bar graph uses the height of bars to show a relationship, where each bar represents a certain entity, like a country or a group of people. The bars on a bar graph can also be divided into segments to show subgroups.

Any graph is a single visual perspective on a subject. The impression it leaves will be based on many choices, such as what data or time frame is included, how data or groups are divided up, the relative size of vertical and horizontal axes, whether the scale used on a vertical starts at zero. Thus, any graph should be regarded somewhat skeptically, remembering that the underlying relationship can be open to different interpretations.

## **Review Questions**

### **Exercise:**

#### **Problem:**

Name three kinds of graphs and briefly state when is most appropriate to use each type of graph.

### **Exercise:**

**Problem:** What is slope on a line graph?

### **Exercise:**

**Problem:** What do the slices of a pie chart represent?

### **Exercise:**

**Problem:** Why is a bar chart the best way to illustrate comparisons?

### **Exercise:**

#### **Problem:**

How does the appearance of positive slope differ from negative slope and from zero slope?

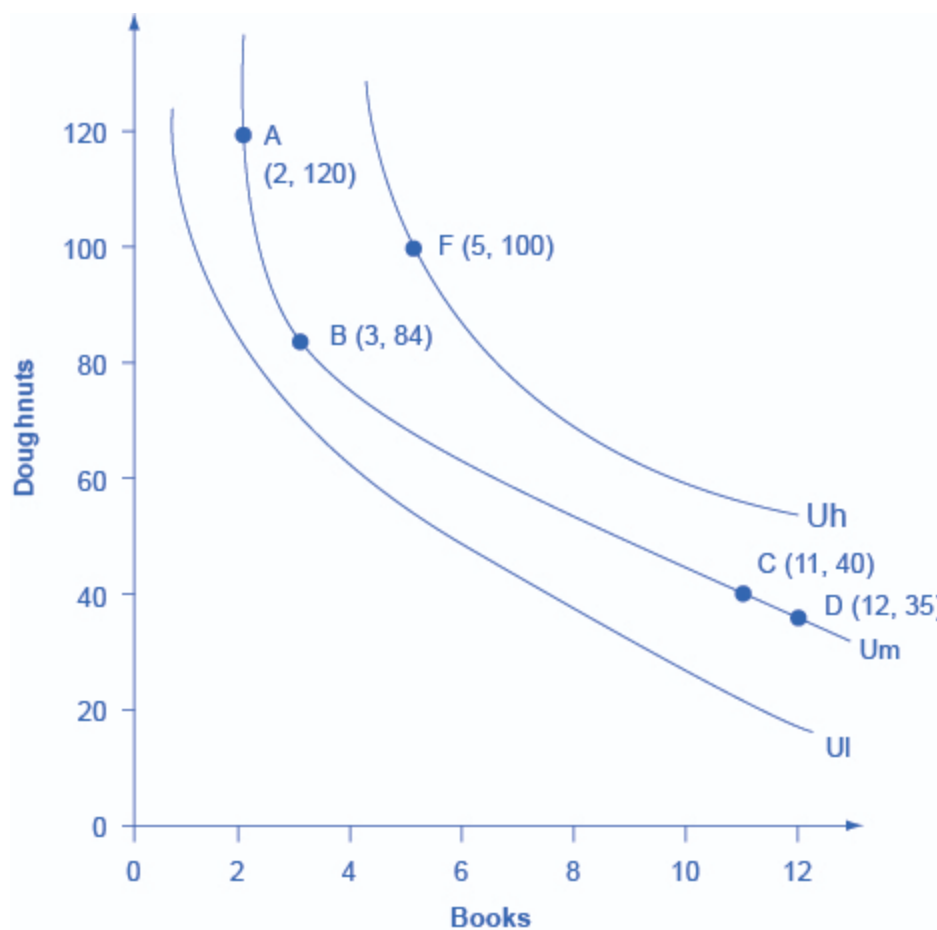
## Indifference Curves

Economists use a vocabulary of maximizing utility to describe people's preferences. In [Consumer Choices](#), the level of utility that a person receives is described in numerical terms. This appendix presents an alternative approach to describing personal preferences, called indifference curves, which avoids any need for using numbers to measure utility. By setting aside the assumption of putting a numerical valuation on utility—an assumption that many students and economists find uncomfortably unrealistic—the indifference curve framework helps to clarify the logic of the underlying model.

### What Is an Indifference Curve?

People cannot really put a numerical value on their level of satisfaction. However, they can, and do, identify what choices would give them more, or less, or the same amount of satisfaction. An indifference curve shows combinations of goods that provide an equal level of utility or satisfaction. For example, [\[link\]](#) presents three indifference curves that represent Lilly's preferences for the tradeoffs that she faces in her two main relaxation activities: eating doughnuts and reading paperback books. Each indifference curve ( $U_l$ ,  $U_m$ , and  $U_h$ ) represents one level of utility. First we will explore the meaning of one particular indifference curve and then we will look at the indifference curves as a group.

#### Lilly's Indifference Curves



Lilly would receive equal utility from all points on a given indifference curve. Any points on the highest indifference curve  $U_h$ , like F, provide greater utility than any points like A, B, C, and D on the middle indifference curve  $U_m$ . Similarly, any points on the middle indifference curve  $U_m$  provide greater utility than any points on the lowest indifference curve  $U_l$ .

### The Shape of an Indifference Curve

The indifference curve  $U_m$  has four points labeled on it: A, B, C, and D. Since an indifference curve represents a set of choices that have the same level of utility, Lilly must receive an equal amount of utility, judged according to her personal preferences, from two books and 120 doughnuts

(point A), from three books and 84 doughnuts (point B) from 11 books and 40 doughnuts (point C) or from 12 books and 35 doughnuts (point D). She would also receive the same utility from any of the unlabeled intermediate points along this indifference curve.

Indifference curves have a roughly similar shape in two ways: 1) they are downward sloping from left to right; 2) they are convex with respect to the origin. In other words, they are steeper on the left and flatter on the right. The downward slope of the indifference curve means that Lilly must trade off less of one good to get more of the other, while holding utility constant. For example, points A and B sit on the same indifference curve  $U_m$ , which means that they provide Lilly with the same level of utility. Thus, the marginal utility that Lilly would gain from, say, increasing her consumption of books from two to three must be equal to the marginal utility that she would lose if her consumption of doughnuts was cut from 120 to 84—so that her overall utility remains unchanged between points A and B. Indeed, the slope along an indifference curve is referred to as the marginal rate of substitution, which is the rate at which a person is willing to trade one good for another so that utility will remain the same.

Indifference curves like  $U_m$  are steeper on the left and flatter on the right. The reason behind this shape involves diminishing marginal utility—the notion that as a person consumes more of a good, the marginal utility from each additional unit becomes lower. Compare two different choices between points that all provide Lilly an equal amount of utility along the indifference curve  $U_m$ : the choice between A and B, and between C and D. In both choices, Lilly consumes one more book, but between A and B her consumption of doughnuts falls by 36 (from 120 to 84) and between C and D it falls by only five (from 40 to 35). The reason for this difference is that points A and C are different starting points, and thus have different implications for marginal utility. At point A, Lilly has few books and many doughnuts. Thus, her marginal utility from an extra book will be relatively high while the marginal utility of additional doughnuts is relatively low—so on the margin, it will take a relatively large number of doughnuts to offset the utility from the marginal book. At point C, however, Lilly has many books and few doughnuts. From this starting point, her marginal utility gained from extra books will be relatively low, while the marginal utility

lost from additional doughnuts would be relatively high—so on the margin, it will take a relatively smaller number of doughnuts to offset the change of one marginal book. In short, the slope of the indifference curve changes because the marginal rate of substitution—that is, the quantity of one good that would be traded for the other good to keep utility constant—also changes, as a result of diminishing marginal utility of both goods.

## **The Field of Indifference Curves**

Each indifference curve represents the choices that provide a single level of utility. Every level of utility will have its own indifference curve. Thus, Lilly's preferences will include an infinite number of indifference curves lying nestled together on the diagram—even though only three of the indifference curves, representing three levels of utility, appear on [\[link\]](#). In other words, an infinite number of indifference curves are not drawn on this diagram—but you should remember that they exist.

Higher indifference curves represent a greater level of utility than lower ones. In [\[link\]](#), indifference curve  $U_l$  can be thought of as a “low” level of utility, while  $U_m$  is a “medium” level of utility and  $U_h$  is a “high” level of utility. All of the choices on indifference curve  $U_h$  are preferred to all of the choices on indifference curve  $U_m$ , which in turn are preferred to all of the choices on  $U_l$ .

To understand why higher indifference curves are preferred to lower ones, compare point B on indifference curve  $U_m$  to point F on indifference curve  $U_h$ . Point F has greater consumption of both books (five to three) and doughnuts (100 to 84), so point F is clearly preferable to point B. Given the definition of an indifference curve—that all the points on the curve have the same level of utility—if point F on indifference curve  $U_h$  is preferred to point B on indifference curve  $U_m$ , then it must be true that all points on indifference curve  $U_h$  have a higher level of utility than all points on  $U_m$ . More generally, for any point on a lower indifference curve, like  $U_l$ , you can identify a point on a higher indifference curve like  $U_m$  or  $U_h$  that has a higher consumption of both goods. Since one point on the higher indifference curve is preferred to one point on the lower curve, and since all the points on a given indifference curve have the same level of utility, it

must be true that all points on higher indifference curves have greater utility than all points on lower indifference curves.

These arguments about the shapes of indifference curves and about higher or lower levels of utility do not require any numerical estimates of utility, either by the individual or by anyone else. They are only based on the assumptions that when people have less of one good they need more of another good to make up for it, if they are keeping the same level of utility, and that as people have more of a good, the marginal utility they receive from additional units of that good will diminish. Given these gentle assumptions, a field of indifference curves can be mapped out to describe the preferences of any individual.

### **The Individuality of Indifference Curves**

Each person determines their own preferences and utility. Thus, while indifference curves have the same general shape—they slope down, and the slope is steeper on the left and flatter on the right—the specific shape of indifference curves can be different for every person. [\[link\]](#), for example, applies only to Lilly’s preferences. Indifference curves for other people would probably travel through different points.

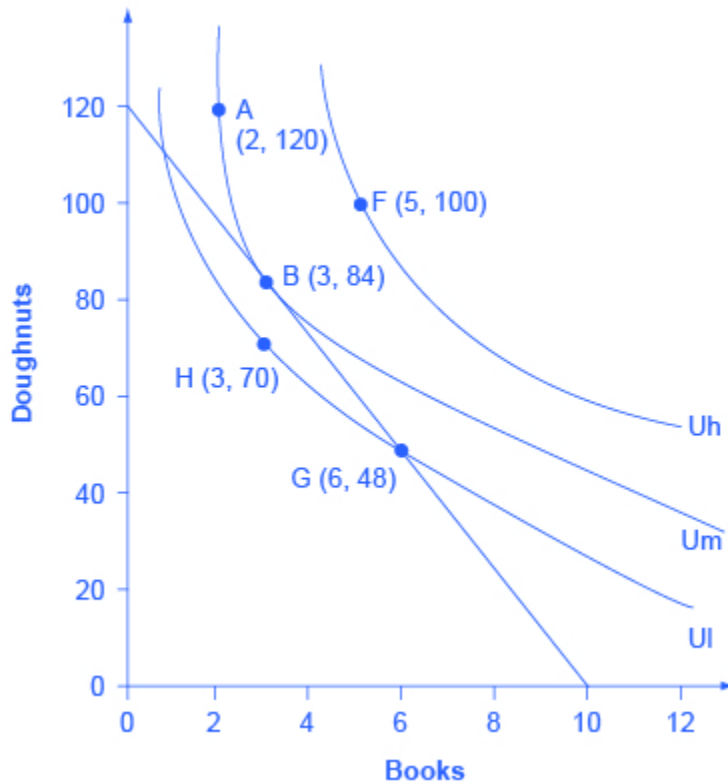
### **Utility-Maximizing with Indifference Curves**

People seek the highest level of utility, which means that they wish to be on the highest possible indifference curve. However, people are limited by their budget constraints, which show what tradeoffs are actually possible.

### **Maximizing Utility at the Highest Indifference Curve**

Return to the situation of Lilly’s choice between paperback books and doughnuts. Say that books cost \$6, doughnuts are 50 cents each, and that Lilly has \$60 to spend. This information provides the basis for the budget line shown in [\[link\]](#). Along with the budget line are shown the three indifference curves from [\[link\]](#). What is Lilly’s utility-maximizing choice? Several possibilities are identified in the diagram.

**Indifference Curves and a Budget Constraint**



Lilly's preferences are shown by the indifference curves. Lilly's budget constraint, given the prices of books and doughnuts and her income, is shown by the straight line. Lilly's optimal choice will be point B, where the budget line is tangent to the indifference curve  $U_m$ . Lilly would have more utility at a point like F on the higher indifference curve  $U_h$ , but the budget line does not touch the higher indifference curve  $U_h$  at any point, so she cannot afford this choice. A choice like G is affordable to Lilly, but it lies on indifference curve  $U_l$  and thus provides less utility than choice B, which is on indifference curve  $U_m$ .

The choice of F with five books and 100 doughnuts is highly desirable, since it is on the highest indifference curve  $U_h$  of those shown in the diagram. However, it is not affordable given Lilly's budget constraint. The choice of H with three books and 70 doughnuts on indifference curve  $U_l$  is a wasteful choice, since it is inside Lilly's budget set, and as a utility-maximizer, Lilly will always prefer a choice on the budget constraint itself. Choices B and G are both on the opportunity set. However, choice G of six books and 48 doughnuts is on lower indifference curve  $U_l$  than choice B of three books and 84 doughnuts, which is on the indifference curve  $U_m$ . If Lilly were to start at choice G, and then thought about whether the marginal utility she was deriving from doughnuts and books, she would decide that some additional doughnuts and fewer books would make her happier—which would cause her to move toward her preferred choice B. Given the combination of Lilly's personal preferences, as identified by her indifference curves, and Lilly's opportunity set, which is determined by prices and income, B will be her utility-maximizing choice.

The highest achievable indifference curve touches the opportunity set at a single point of tangency. Since an infinite number of indifference curves exist, even if only a few of them are drawn on any given diagram, there will always exist one indifference curve that touches the budget line at a single point of tangency. All higher indifference curves, like  $U_h$ , will be completely above the budget line and, although the choices on that indifference curve would provide higher utility, they are not affordable given the budget set. All lower indifference curves, like  $U_l$ , will cross the budget line in two separate places. When one indifference curve crosses the budget line in two places, however, there will be another, higher, attainable indifference curve sitting above it that touches the budget line at only one point of tangency.

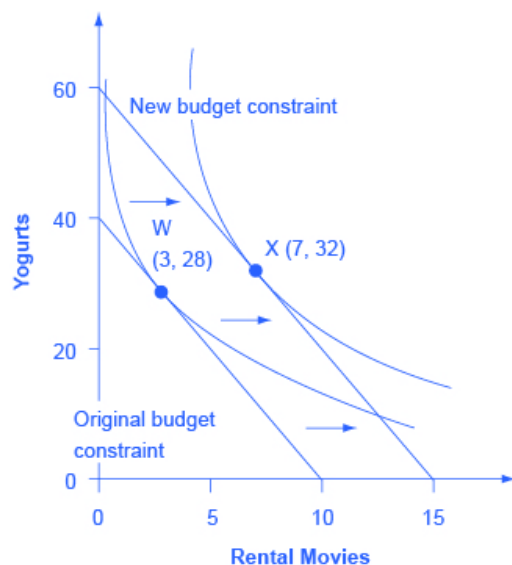
## **Changes in Income**

A rise in income causes the budget constraint to shift to the right. In graphical terms, the new budget constraint will now be tangent to a higher indifference curve, representing a higher level of utility. A reduction in income will cause the budget constraint to shift to the left, which will cause it to be tangent to a lower indifference curve, representing a reduced level

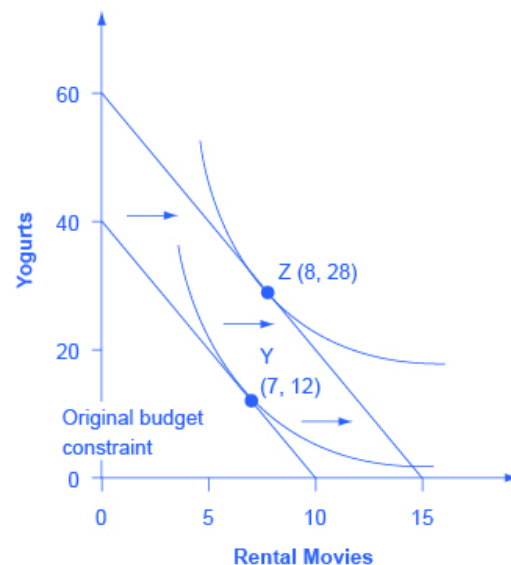
of utility. If income rises by, for example, 50%, exactly how much will a person alter consumption of books and doughnuts? Will consumption of both goods rise by 50%? Or will the quantity of one good rise substantially, while the quantity of the other good rises only a little, or even declines?

Since personal preferences and the shape of indifference curves are different for each individual, the response to changes in income will be different, too. For example, consider the preferences of Manuel and Natasha in [\[link\]](#) (a) and [\[link\]](#) (b). They each start with an identical income of \$40, which they spend on yogurts that cost \$1 and rental movies that cost \$4. Thus, they face identical budget constraints. However, based on Manuel's preferences, as revealed by his indifference curves, his utility-maximizing choice on the original budget set occurs where his opportunity set is tangent to the highest possible indifference curve at W, with three movies and 28 yogurts, while Natasha's utility-maximizing choice on the original budget set at Y will be seven movies and 12 yogurts.

### Manuel and Natasha's Indifference Curves



(a) Manuel's reaction to more income



(b) Natasha's reaction to more income

Manuel and Natasha originally face the same budget constraints; that is, same prices and same income. However, the indifference curves that illustrate their preferences are not the same. (a) Manuel's original choice at W involves more yogurt and more movies, and he reacts to the higher income by mainly increasing consumption of movies at X.

(b) Conversely, Natasha's original choice (Y) involves relatively more movies, but she reacts to the higher income by choosing relatively more yogurts. Even when budget constraints are the same, personal preferences lead to different original choices and to different reactions in response to a change in income.

Now, say that income rises to \$60 for both Manuel and Natasha, so their budget constraints shift to the right. As shown in [\[link\]](#) (a), Manuel's new utility maximizing choice at X will be seven movies and 32 yogurts—that is, Manuel will choose to spend most of the extra income on movies. Natasha's new utility maximizing choice at Z will be eight movies and 28 yogurts—that is, she will choose to spend most of the extra income on yogurt. In this way, the indifference curve approach allows for a range of possible responses. However, if both goods are normal goods, then the typical response to a higher level of income will be to purchase more of them—although exactly how much more is a matter of personal preference. If one of the goods is an inferior good, the response to a higher level of income will be to purchase less of it.

## **Responses to Price Changes: Substitution and Income Effects**

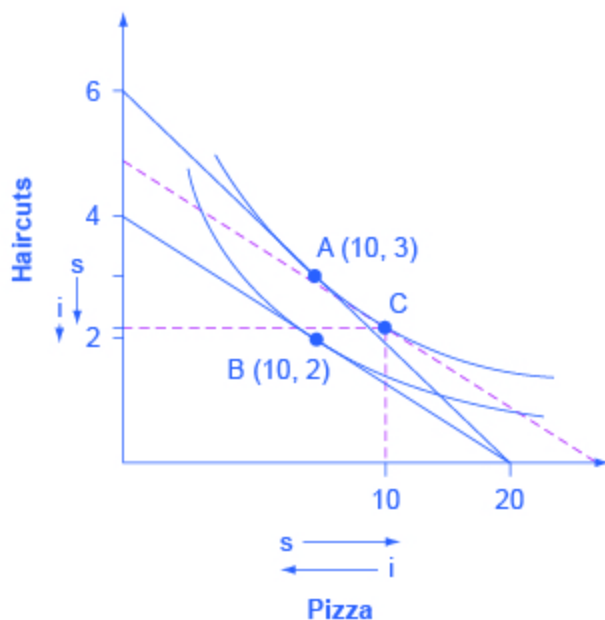
A higher price for a good will cause the budget constraint to shift to the left, so that it is tangent to a lower indifference curve representing a reduced level of utility. Conversely, a lower price for a good will cause the opportunity set to shift to the right, so that it is tangent to a higher indifference curve representing an increased level of utility. Exactly how much a change in price will lead to the quantity demanded of each good will depend on personal preferences.

Anyone who faces a change in price will experience two interlinked motivations: a substitution effect and an income effect. The substitution effect is that when a good becomes more expensive, people seek out substitutes. If oranges become more expensive, fruit-lovers scale back on oranges and eat more apples, grapefruit, or raisins. Conversely, when a good becomes cheaper, people substitute toward consuming more. If

oranges get cheaper, people fire up their juicing machines and ease off on other fruits and foods. The income effect refers to how a change in the price of a good alters the effective buying power of one's income. If the price of a good that you have been buying falls, then in effect your buying power has risen—you are able to purchase more goods. Conversely, if the price of a good that you have been buying rises, then the buying power of a given amount of income is diminished. (One common source of confusion is that the “income effect” does not refer to a change in actual income. Instead, it refers to the situation in which the price of a good changes, and thus the quantities of goods that can be purchased with a fixed amount of income change. It might be more accurate to call the “income effect” a “buying power effect,” but the “income effect” terminology has been used for decades, and it is not going to change during this economics course.) Whenever a price changes, consumers feel the pull of both substitution and income effects at the same time.

Using indifference curves, you can illustrate the substitution and income effects on a graph. In [\[link\]](#), Ogden faces a choice between two goods: haircuts or personal pizzas. Haircuts cost \$20, personal pizzas cost \$6, and he has \$120 to spend.

### Substitution and Income Effects



The original choice is A, the point of

tangency between the original budget constraint and indifference curve. The new choice is B, the point of tangency between the new budget constraint and the lower indifference curve. Point C is the tangency between the dashed line, where the slope shows the new higher price of haircuts, and the original indifference curve. The substitution effect is the shift from A to C, which means getting fewer haircuts and more pizza. The income effect is the shift from C to B; that is, the reduction in buying power that causes a shift from the higher indifference curve to the lower indifference curve, with relative prices remaining unchanged. The income effect results in less consumed of both goods. Both substitution and income effects cause fewer haircuts to be consumed. For pizza, in this case, the substitution effect and income effect cancel out, leading to the same amount of pizza consumed.

The price of haircuts rises to \$30. Ogden starts at choice A on the higher opportunity set and the higher indifference curve. After the price of pizza increases, he chooses B on the lower opportunity set and the lower indifference curve. Point B with two haircuts and 10 personal pizzas is immediately below point A with three haircuts and 10 personal pizzas, showing that Ogden reacted to a higher price of haircuts by cutting back only on haircuts, while leaving his consumption of pizza unchanged.

The dashed line in the diagram, and point C, are used to separate the substitution effect and the income effect. To understand their function, start by thinking about the substitution effect with this question: How would Ogden change his consumption if the relative prices of the two goods

changed, but this change in relative prices did not affect his utility? The slope of the budget constraint is determined by the relative price of the two goods; thus, the slope of the original budget line is determined by the original relative prices, while the slope of the new budget line is determined by the new relative prices. With this thought in mind, the dashed line is a graphical tool inserted in a specific way: It is inserted so that it is parallel with the new budget constraint, so it reflects the new relative prices, but it is tangent to the original indifference curve, so it reflects the original level of utility or buying power.

Thus, the movement from the original choice (A) to point C is a substitution effect; it shows the choice that Ogden would make if relative prices shifted (as shown by the different slope between the original budget set and the dashed line) but if buying power did not shift (as shown by being tangent to the original indifference curve). The substitution effect will encourage people to shift away from the good which has become relatively more expensive—in Ogden's case, the haircuts on the vertical axis—and toward the good which has become relatively less expensive—in this case, the pizza on the vertical axis. The two arrows labeled with “s” for “substitution effect,” one on each axis, show the direction of this movement.

The income effect is the movement from point C to B, which shows how Ogden reacts to a reduction in his buying power from the higher indifference curve to the lower indifference curve, but holding constant the relative prices (because the dashed line has the same slope as the new budget constraint). In this case, where the price of one good increases, buying power is reduced, so the income effect means that consumption of both goods should fall (if they are both normal goods, which it is reasonable to assume unless there is reason to believe otherwise). The two arrows labeled with “i” for “income effect,” one on each axis, show the direction of this income effect movement.

Now, put the substitution and income effects together. When the price of pizza increased, Ogden consumed less of it, for two reasons shown in the exhibit: the substitution effect of the higher price led him to consume less and the income effect of the higher price also led him to consume less. However, when the price of pizza increased, Ogden consumed the same

quantity of haircuts. The substitution effect of a higher price for pizza meant that haircuts became relatively less expensive (compared to pizza), and this factor, taken alone, would have encouraged Ogden to consume more haircuts. However, the income effect of a higher price for pizza meant that he wished to consume less of both goods, and this factor, taken alone, would have encouraged Ogden to consume fewer haircuts. As shown in [\[link\]](#), in this particular example the substitution effect and income effect on Ogden's consumption of haircuts are offsetting—so he ends up consuming the same quantity of haircuts after the price increase for pizza as before.

The size of these income and substitution effects will differ from person to person, depending on individual preferences. For example, if Ogden's substitution effect away from pizza and toward haircuts is especially strong, and outweighs the income effect, then a higher price for pizza might lead to increased consumption of haircuts. This case would be drawn on the graph so that the point of tangency between the new budget constraint and the relevant indifference curve occurred below point B and to the right. Conversely, if the substitution effect away from pizza and toward haircuts is not as strong, and the income effect on is relatively stronger, then Ogden will be more likely to react to the higher price of pizza by consuming less of both goods. In this case, his optimal choice after the price change will be above and to the left of choice B on the new budget constraint.

Although the substitution and income effects are often discussed as a sequence of events, it should be remembered that they are twin components of a single cause—a change in price. Although you can analyze them separately, the two effects are always proceeding hand in hand, happening at the same time.

## **Indifference Curves with Labor-Leisure and Intertemporal Choices**

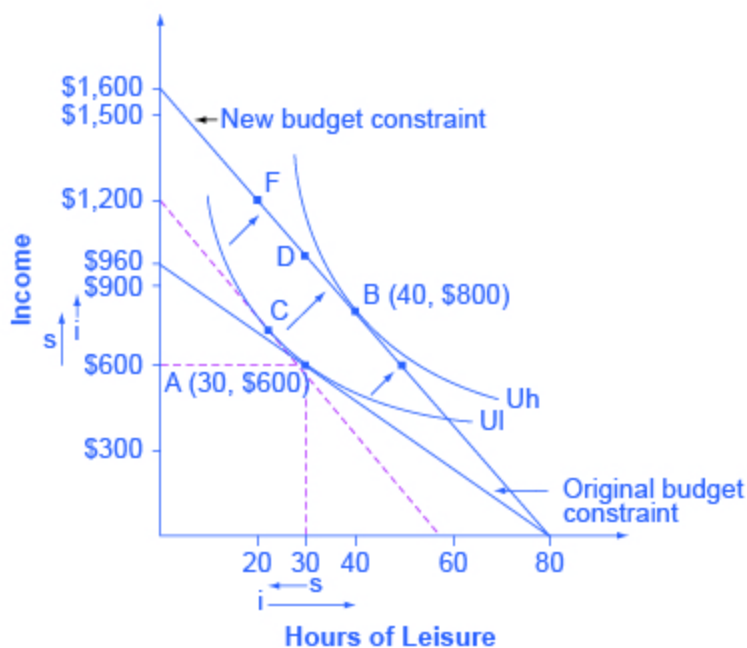
The concept of an indifference curve applies to tradeoffs in any household choice, including the labor-leisure choice or the intertemporal choice between present and future consumption. In the labor-leisure choice, each indifference curve shows the combinations of leisure and income that provide a certain level of utility. In an intertemporal choice, each

indifference curve shows the combinations of present and future consumption that provide a certain level of utility. The general shapes of the indifference curves—downward sloping, steeper on the left and flatter on the right—also remain the same.

### A Labor-Leisure Example

Petunia is working at a job that pays \$12 per hour but she gets a raise to \$20 per hour. After family responsibilities and sleep, she has 80 hours per week available for work or leisure. As shown in [\[link\]](#), the highest level of utility for Petunia, on her original budget constraint, is at choice A, where it is tangent to the lower indifference curve (U<sub>l</sub>). Point A has 30 hours of leisure and thus 50 hours per week of work, with income of \$600 per week (that is, 50 hours of work at \$12 per hour). Petunia then gets a raise to \$20 per hour, which shifts her budget constraint to the right. Her new utility-maximizing choice occurs where the new budget constraint is tangent to the higher indifference curve U<sub>h</sub>. At B, Petunia has 40 hours of leisure per week and works 40 hours, with income of \$800 per week (that is, 40 hours of work at \$20 per hour).

Effects of a Change in Petunia's Wage



Petunia starts at choice A, the tangency

between her original budget constraint and the lower indifference curve  $U_l$ . The wage increase shifts her budget constraint to the right, so that she can now choose B on indifference curve  $U_h$ . The substitution effect is the movement from A to C. In this case, the substitution effect would lead Petunia to choose less leisure, which is relatively more expensive, and more income, which is relatively cheaper to earn. The income effect is the movement from C to B. The income effect in this example leads to greater consumption of both goods. Overall, in this example, income rises because of both substitution and income effects. However, leisure declines because of the substitution effect but increases because of the income effect—leading, in Petunia’s case, to an overall increase in the quantity of leisure consumed.

Substitution and income effects provide a vocabulary for discussing how Petunia reacts to a higher hourly wage. The dashed line serves as the tool for separating the two effects on the graph.

The substitution effect tells how Petunia would have changed her hours of work if her wage had risen, so that income was relatively cheaper to earn and leisure was relatively more expensive, but if she had remained at the same level of utility. The slope of the budget constraint in a labor-leisure diagram is determined by the wage rate. Thus, the dashed line is carefully inserted with the slope of the new opportunity set, reflecting the labor-leisure tradeoff of the new wage rate, but tangent to the original indifference curve, showing the same level of utility or “buying power.” The shift from original choice A to point C, which is the point of tangency between the

original indifference curve and the dashed line, shows that because of the higher wage, Petunia will want to consume less leisure and more income. The “s” arrows on the horizontal and vertical axes of [\[link\]](#) show the substitution effect on leisure and on income.

The income effect is that the higher wage, by shifting the labor-leisure budget constraint to the right, makes it possible for Petunia to reach a higher level of utility. The income effect is the movement from point C to point B; that is, it shows how Petunia’s behavior would change in response to a higher level of utility or “buying power,” with the wage rate remaining the same (as shown by the dashed line being parallel to the new budget constraint). The income effect, encouraging Petunia to consume both more leisure and more income, is drawn with arrows on the horizontal and vertical axis of [\[link\]](#).

Putting these effects together, Petunia responds to the higher wage by moving from choice A to choice B. This movement involves choosing more income, both because the substitution effect of higher wages has made income relatively cheaper or easier to earn, and because the income effect of higher wages has made it possible to have more income and more leisure. Her movement from A to B also involves choosing more leisure because, according to Petunia’s preferences, the income effect that encourages choosing more leisure is stronger than the substitution effect that encourages choosing less leisure.

[\[link\]](#) represents only Petunia’s preferences. Other people might make other choices. For example, a person whose substitution and income effects on leisure exactly counterbalanced each other might react to a higher wage with a choice like D, exactly above the original choice A, which means taking all of the benefit of the higher wages in the form of income while working the same number of hours. Yet another person, whose substitution effect on leisure outweighed the income effect, might react to a higher wage by making a choice like F, where the response to higher wages is to work more hours and earn much more income. To represent these different preferences, you could easily draw the indifference curve  $U_h$  to be tangent to the new budget constraint at D or F, rather than at B.

## **An Intertemporal Choice Example**

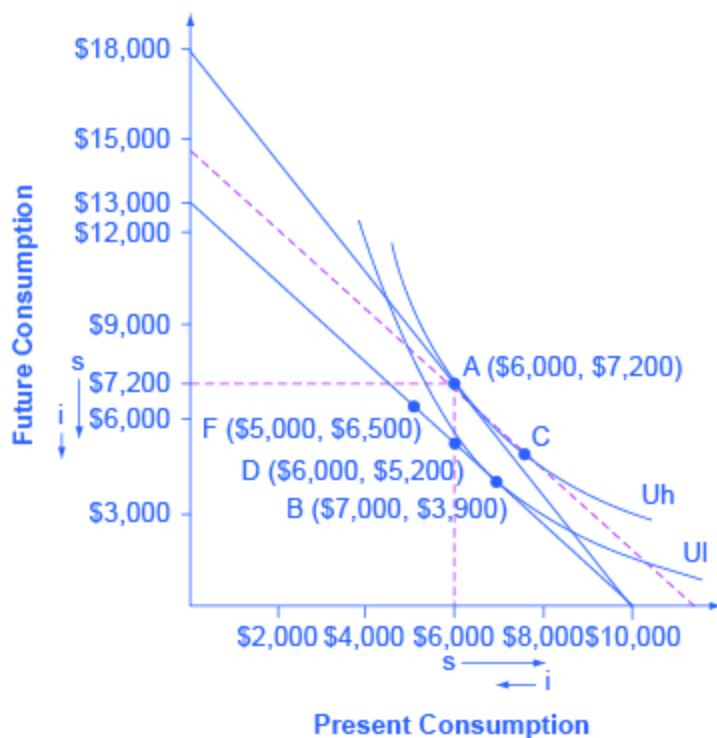
Quentin has saved up \$10,000. He is thinking about spending some or all of it on a vacation in the present, and then will save the rest for another big vacation five years from now. Over those five years, he expects to earn a total 80% rate of return. [\[link\]](#) shows Quentin's budget constraint and his indifference curves between present consumption and future consumption. The highest level of utility that Quentin can achieve at his original intertemporal budget constraint occurs at point A, where he is consuming \$6,000, saving \$4,000 for the future, and expecting with the accumulated interest to have \$7,200 for future consumption (that is, \$4,000 in current financial savings plus the 80% rate of return).

However, Quentin has just realized that his expected rate of return was unrealistically high. A more realistic expectation is that over five years he can earn a total return of 30%. In effect, his intertemporal budget constraint has pivoted to the left, so that his original utility-maximizing choice is no longer available. Will Quentin react to the lower rate of return by saving more, or less, or the same amount? Again, the language of substitution and income effects provides a framework for thinking about the motivations behind various choices. The dashed line, which is a graphical tool to separate the substitution and income effect, is carefully inserted with the same slope as the new opportunity set, so that it reflects the changed rate of return, but it is tangent to the original indifference curve, so that it shows no change in utility or "buying power."

The substitution effect tells how Quentin would have altered his consumption because the lower rate of return makes future consumption relatively more expensive and present consumption relatively cheaper. The movement from the original choice A to point C shows how Quentin substitutes toward more present consumption and less future consumption in response to the lower interest rate, with no change in utility. The substitution arrows on the horizontal and vertical axes of [\[link\]](#) show the direction of the substitution effect motivation. The substitution effect suggests that, because of the lower interest rate, Quentin should consume more in the present and less in the future.

Quentin also has an income effect motivation. The lower rate of return shifts the budget constraint to the left, which means that Quentin's utility or

“buying power” is reduced. The income effect (assuming normal goods) encourages less of both present and future consumption. The impact of the income effect on reducing present and future consumption in this example is shown with “i” arrows on the horizontal and vertical axis of [\[link\]](#). Indifference Curve and an Intertemporal Budget Constraint



The original choice is A, at the tangency between the original budget constraint and the original indifference curve  $U_h$ . The dashed line is drawn parallel to the new budget set, so that its slope reflects the lower rate of return, but is tangent to the original indifference curve. The movement from A to C is the substitution effect: in this case, future consumption has become relatively more expensive, and present consumption has become relatively cheaper. The income effect is the shift from C to B; that is, the reduction in utility or “buying power” that causes a move to a

lower indifference curve  $U_1$ , but with the relative price the same. It means less present and less future consumption. In the move from A to B, the substitution effect on present consumption is greater than the income effect, so the overall result is more present consumption. Notice that the lower indifference curve could have been drawn tangent to the lower budget constraint point D or point F, depending on personal preferences.

Taking both effects together, the substitution effect is encouraging Quentin toward more present and less future consumption, because present consumption is relatively cheaper, while the income effect is encouraging him to less present and less future consumption, because the lower interest rate is pushing him to a lower level of utility. For Quentin's personal preferences, the substitution effect is stronger so that, overall, he reacts to the lower rate of return with more present consumption and less savings at choice B. However, other people might have different preferences. They might react to a lower rate of return by choosing the same level of present consumption and savings at choice D, or by choosing less present consumption and more savings at a point like F. For these other sets of preferences, the income effect of a lower rate of return on present consumption would be relatively stronger, while the substitution effect would be relatively weaker.

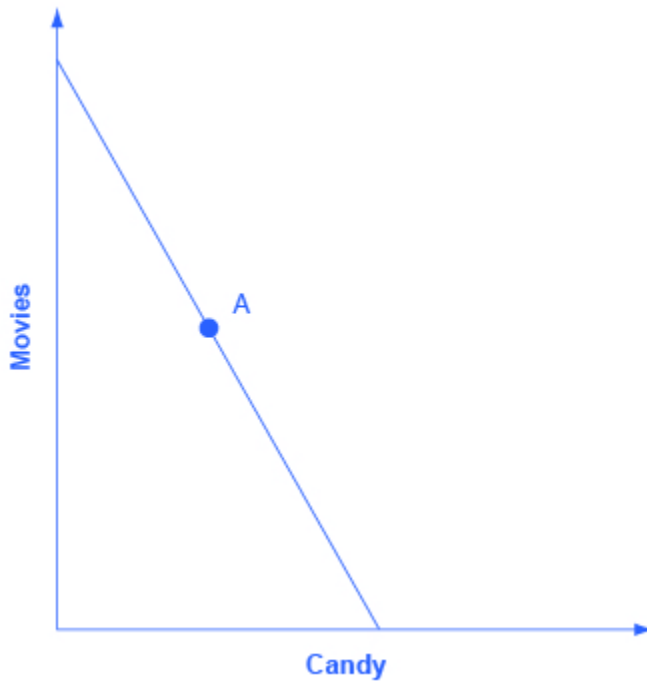
## **Sketching Substitution and Income Effects**

Indifference curves provide an analytical tool for looking at all the choices that provide a single level of utility. They eliminate any need for placing numerical values on utility and help to illuminate the process of making utility-maximizing decisions. They also provide the basis for a more detailed investigation of the complementary motivations that arise in

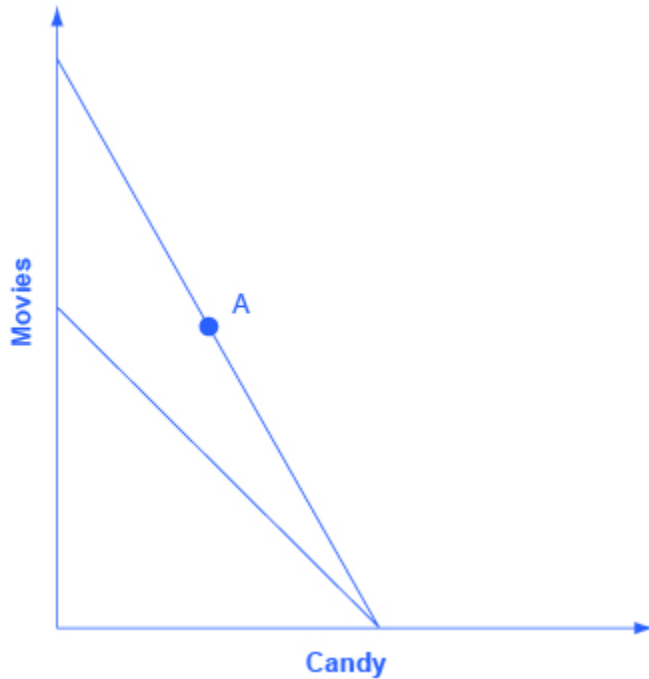
response to a change in a price, wage or rate of return—namely, the substitution and income effects.

If you are finding it a little tricky to sketch diagrams that show substitution and income effects so that the points of tangency all come out correctly, it may be useful to follow this procedure.

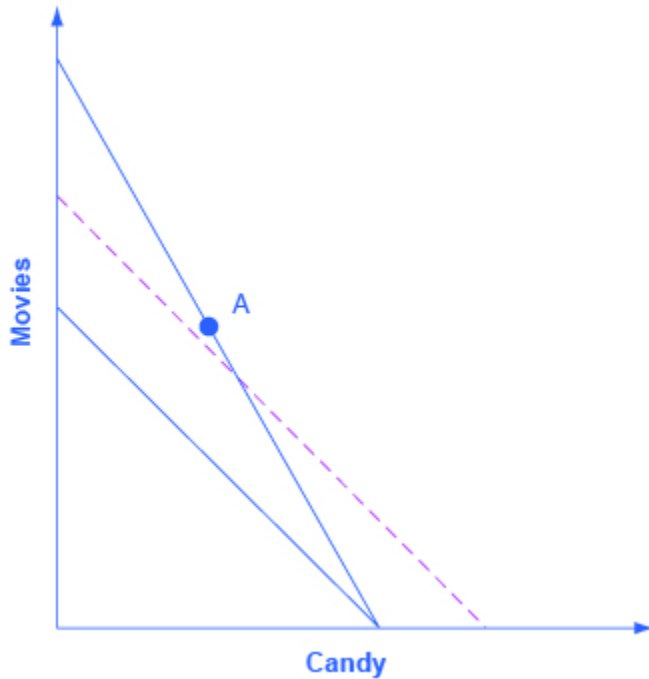
Step 1. Begin with a budget constraint showing the choice between two goods, which this example will call “candy” and “movies.” Choose a point A which will be the optimal choice, where the indifference curve will be tangent—but it is often easier not to draw in the indifference curve just yet. See [\[link\]](#).



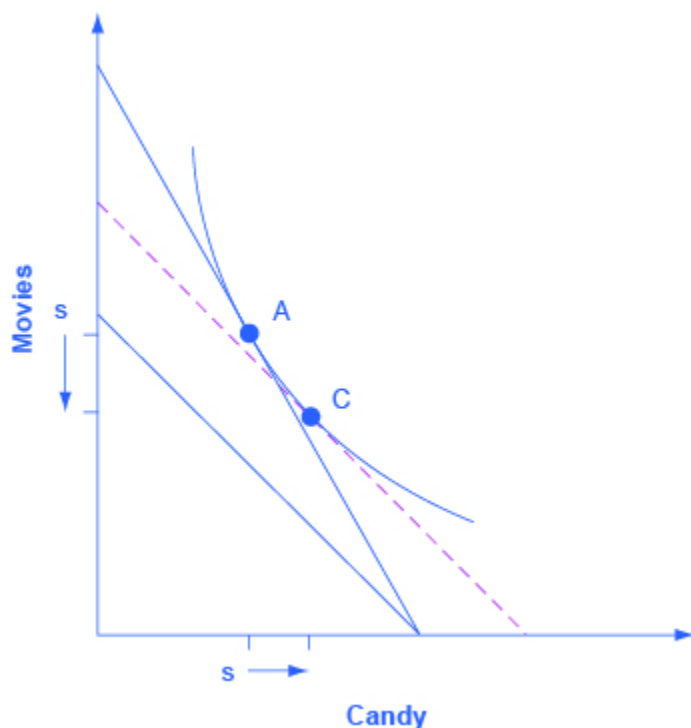
Step 2. Now the price of movies changes: let’s say that it rises. That shifts the budget set inward. You know that the higher price will push the decision-maker down to a lower level of utility, represented by a lower indifference curve. But at this stage, draw only the new budget set. See [\[link\]](#).



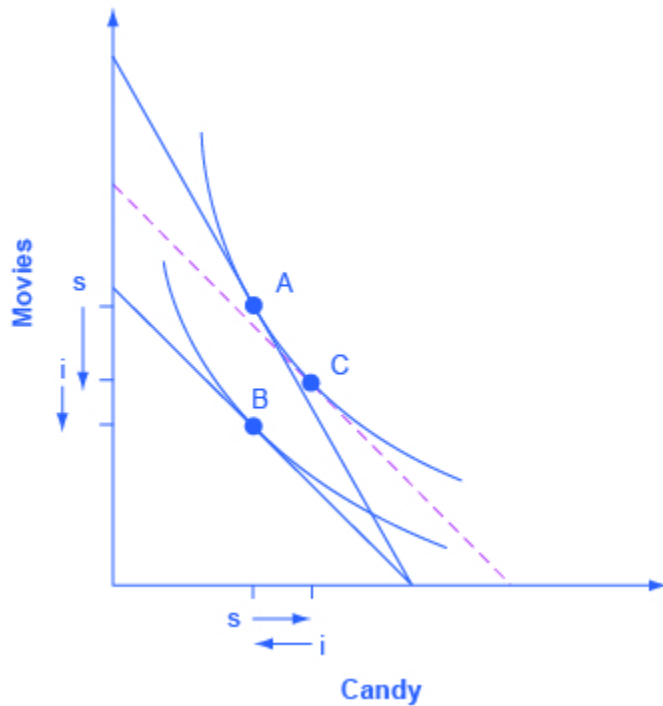
Step 3. The key tool in distinguishing between substitution and income effects is to insert a dashed line, parallel to the new budget line. This line is a graphical tool that allows you to distinguish between the two changes: (1) the effect on consumption of the two goods of the shift in prices—with the level of utility remaining unchanged—which is the substitution effect; and (2) the effect on consumption of the two goods of shifting from one indifference curve to the other—with relative prices staying unchanged—which is the income effect. The dashed line is inserted in this step. The trick is to have the dashed line travel close to the original choice A, but not directly through point A. See [\[link\]](#).



Step 4. Now, draw the original indifference curve, so that it is tangent to both point A on the original budget line and to a point C on the dashed line. Many students find it easiest to first select the tangency point C where the original indifference curve touches the dashed line, and then to draw the original indifference curve through A and C. The substitution effect is illustrated by the movement along the original indifference curve as prices change but the level of utility holds constant, from A to C. As expected, the substitution effect leads to less consumed of the good that is relatively more expensive, as shown by the “s” (substitution) arrow on the vertical axis, and more consumed of the good that is relatively less expensive, as shown by the “s” arrow on the horizontal axis. See [\[link\]](#).



Step 5. With the substitution effect in place, now choose utility-maximizing point B on the new opportunity set. When you choose point B, think about whether you wish the substitution or the income effect to have a larger impact on the good (in this case, candy) on the horizontal axis. If you choose point B to be directly in a vertical line with point A (as is illustrated here), then the income effect will be exactly offsetting the substitution effect on the horizontal axis. If you insert point B so that it lies a little to right of the original point A, then the substitution effect will exceed the income effect. If you insert point B so that it lies a little to the left of point A, then the income effect will exceed the substitution effect. The income effect is the movement from C to B, showing how choices shifted as a result of the decline in buying power and the movement between two levels of utility, with relative prices remaining the same. With normal goods, the negative income effect means less consumed of each good, as shown by the direction of the “i” (income effect) arrows on the vertical and horizontal axes. See [\[link\]](#).



In sketching substitution and income effect diagrams, you may wish to practice some of the following variations: (1) Price falls instead of a rising; (2) The price change affects the good on either the vertical or the horizontal axis; (3) Sketch these diagrams so that the substitution effect exceeds the income effect; the income effect exceeds the substitution effect; and the two effects are equal.

One final note: The helpful dashed line can be drawn tangent to the new indifference curve, and parallel to the original budget line, rather than tangent to the original indifference curve and parallel to the new budget line. Some students find this approach more intuitively clear. The answers you get about the direction and relative sizes of the substitution and income effects, however, should be the same.

## Key Concepts and Summary

An indifference curve is drawn on a budget constraint diagram that shows the tradeoffs between two goods. All points along a single indifference curve provide the same level of utility. Higher indifference curves represent higher levels of utility. Indifference curves slope downward because, if utility is to remain the same at all points along the curve, a reduction in the quantity of the good on the vertical axis must be counterbalanced by an increase in the quantity of the good on the horizontal axis (or vice versa). Indifference curves are steeper on the far left and flatter on the far right, because of diminishing marginal utility.

The utility-maximizing choice along a budget constraint will be the point of tangency where the budget constraint touches an indifference curve at a single point. A change in the price of any good has two effects: a substitution effect and an income effect. The substitution effect motivation encourages a utility-maximizer to buy less of what is relatively more expensive and more of what is relatively cheaper. The income effect motivation encourages a utility-maximizer to buy more of both goods if utility rises or less of both goods if utility falls (if they are both normal goods).

In a labor-leisure choice, every wage change has a substitution and an income effect. The substitution effect of a wage increase is to choose more income, since it is cheaper to earn, and less leisure, since its opportunity cost has increased. The income effect of a wage increase is to choose more of leisure and income, since they are both normal goods. The substitution and income effects of a wage decrease would reverse these directions.

In an intertemporal consumption choice, every interest rate change has a substitution and an income effect. The substitution effect of an interest rate increase is to choose more future consumption, since it is now cheaper to earn future consumption and less present consumption (more savings), since the opportunity cost of present consumption in terms of what is being given up in the future has increased. The income effect of an interest rate increase is to choose more of both present and future consumption, since they are both normal goods. The substitution and income effects of an interest rate decrease would reverse these directions.

## Review Questions

### Exercise:

**Problem:** What point is preferred along an indifference curve?

### Exercise:

**Problem:** Why do indifference curves slope down?

### Exercise:

**Problem:**

Why are indifference curves steep on the left and flatter on the right?

### Exercise:

**Problem:** How many indifference curves does a person have?

### Exercise:

**Problem:**

How can you tell which indifference curves represent higher or lower levels of utility?

### Exercise:

**Problem:** What is a substitution effect?

### Exercise:

**Problem:** What is an income effect?

### Exercise:

**Problem:**

Does the “income effect” involve a change in income? Explain.

### Exercise:

**Problem:**

Does a change in price have both an income effect and a substitution effect? Does a change in income have both an income effect and a substitution effect?

**Exercise:**

**Problem:**

Would you expect, in some cases, to see only an income effect or only a substitution effect? Explain.

**Exercise:**

**Problem:** Which is larger, the income effect or the substitution effect?

## Present Discounted Value

As explained in [Financial Markets](#), the prices of stocks and bonds depend on future events. The price of a bond depends on the future payments that the bond is expected to make, including both payments of interest and the repayment of the face value of the bond. The price of a stock depends on the expected future profits earned by the firm. The concept of a present discounted value (PDV), which is defined as the amount you should be willing to pay in the present for a stream of expected future payments, can be used to calculate appropriate prices for stocks and bonds. To place a present discounted value on a future payment, think about what amount of money you would need to have in the present to equal a certain amount in the future. This calculation will require an interest rate. For example, if the interest rate is 10%, then a payment of \$110 a year from now will have a present discounted value of \$100—that is, you could take \$100 in the present and have \$110 in the future. We will first show how to apply the idea of present discounted value to a stock and then we will show how to apply it to a bond.

### Applying Present Discounted Value to a Stock

Consider the case of Babble, Inc., a company that offers speaking lessons. For the sake of simplicity, say that the founder of Babble is 63 years old and plans to retire in two years, at which point the company will be disbanded. The company is selling 200 shares of stock and profits are expected to be \$15 million right away, in the present, \$20 million one year from now, and \$25 million two years from now. All profits will be paid out as dividends to shareholders as they occur. Given this information, what will an investor pay for a share of stock in this company?

A financial investor, thinking about what future payments are worth in the present, will need to choose an interest rate. This interest rate will reflect the rate of return on other available financial investment opportunities, which is the opportunity cost of investing financial capital, and also a risk premium (that is, using a higher interest rate than the rates available elsewhere if this investment appears especially risky). In this example, say that the financial

investor decides that appropriate interest rate to value these future payments is 15%.

[\[link\]](#) shows how to calculate the present discounted value of the future profits. For each time period, when a benefit is going to be received, apply the formula:

**Equation:**

$$\text{Present discounted value} = \frac{\text{Future value received years in the future}}{(1 + \text{Interest rate})^{\text{numbers of years } t}}$$

Payments from Firm	Present Value
\$15 million in present	\$15 million
\$20 million in one year	$\$20 \text{ million} / (1 + 0.15)^1 = \$17.4 \text{ million}$
\$25 million in two years	$\$25 \text{ million} / (1 + 0.15)^2 = \$18.9 \text{ million}$
<i>Total</i>	<i>\$51.3 million</i>

### Calculating Present Discounted Value of a Stock

Next, add up all the present values for the different time periods to get a final answer. The present value calculations ask what the amount in the future is worth in the present, given the 15% interest rate. Notice that a different PDV calculation needs to be done separately for amounts received at different times. Then, divide the PDV of total profits by the number of shares, 200 in this case:  $51.3 \text{ million} / 200 = 0.2565 \text{ million}$ . The price per share should be about \$256,500 per share.

Of course, in the real world expected profits are a best guess, not a hard piece of data. Deciding which interest rate to apply for discounting to the present can be tricky. One needs to take into account both potential capital gains from the future sale of the stock and also dividends that might be paid. Differences of opinion on these issues are exactly why some financial investors want to buy a stock that other people want to sell: they are more optimistic about its future prospects. Conceptually, however, it all comes down to what you are willing to pay in the present for a stream of benefits to be received in the future.

## **Applying Present Discounted Value to a Bond**

A similar calculation works in the case of bonds. [Financial Markets](#) explains that if the interest rate falls after a bond is issued, so that the investor has locked in a higher rate, then that bond will sell for more than its face value. Conversely, if the interest rate rises after a bond is issued, then the investor is locked into a lower rate, and the bond will sell for less than its face value. The present value calculation sharpens this intuition.

Think about a simple two-year bond. It was issued for \$3,000 at an interest rate of 8%. Thus, after the first year, the bond pays interest of 240 (which is  $3,000 \times 8\%$ ). At the end of the second year, the bond pays \$240 in interest, plus the \$3,000 in principle. Calculate how much this bond is worth in the present if the discount rate is 8%. Then, recalculate if interest rates rise and the applicable discount rate is 11%. To carry out these calculations, look at the stream of payments being received from the bond in the future and figure out what they are worth in present discounted value terms. The calculations applying the present value formula are shown in [\[link\]](#).

<b>Stream of Payments (for the 8% interest rate)</b>	<b>Present Value (for the 8% interest rate)</b>	<b>Stream of Payments (for the 11% interest rate)</b>	<b>Present Value (for the 11% interest rate)</b>
\$240 payment after one year	$\$240/(1 + 0.08)^1 = \$222.20$	\$240 payment after one year	$\$240/(1 + 0.11)^1 = \$216.20$
\$3,240 payment after second year	$\$3,240/(1 + 0.08)^2 = \$2,777.80$	\$3,240 payment after second year	$\$3,240/(1 + 0.11)^2 = \$2,629.60$
<i>Total</i>	<i>\$3,000</i>	<i>Total</i>	<i>\$2,845.80</i>

### Computing the Present Discounted Value of a Bond

The first calculation shows that the present value of a \$3,000 bond, issued at 8%, is just \$3,000. After all, that is how much money the borrower is receiving. The calculation confirms that the present value is the same for the lender. The bond is moving money around in time, from those willing to save in the present to those who want to borrow in the present, but the present value of what is received by the borrower is identical to the present value of what will be repaid to the lender.

The second calculation shows what happens if the interest rate rises from 8% to 11%. The actual dollar payments in the first column, as determined by the 8% interest rate, do not change. However, the present value of those payments, now discounted at a higher interest rate, is lower. Even though the future dollar payments that the bond is receiving have not changed, a person who tries to sell the bond will find that the investment's value has fallen.

Again, real-world calculations are often more complex, in part because, not only the interest rate prevailing in the market, but also the riskiness of whether the borrower will repay the loan, will change. In any case, the price

of a bond is always the present value of a stream of future expected payments.

## **Other Applications**

Present discounted value is a widely used analytical tool outside the world of finance. Every time a business thinks about making a physical capital investment, it must compare a set of present costs of making that investment to the present discounted value of future benefits. When government thinks about a proposal to, for example, add safety features to a highway, it must compare costs incurred in the present to benefits received in the future. Some academic disputes over environmental policies, like how much to reduce carbon dioxide emissions because of the risk that they will lead to a warming of global temperatures several decades in the future, turn on how one compares present costs of pollution control with long-run future benefits. Someone who wins the lottery and is scheduled to receive a string of payments over 30 years might be interested in knowing what the present discounted value is of those payments. Whenever a string of costs and benefits stretches from the present into different times in the future, present discounted value becomes an indispensable tool of analysis.